

**Etablierung einer neuen Methode zur gerichteten Sequenzierung im Next Generation
Sequencing von cDNA-Banken**

Von der Fakultät für Lebenswissenschaften
der Technischen Universität Carolo-Wilhelmina zu Braunschweig
zur Erlangung des Grades
eines Doktors der Naturwissenschaften (Dr.rer.nat.)

genehmigte

D i s s e r t a t i o n

von Florian Erich Knaust

aus Berlin

1. Referent: Professor Dr. Stefan Dübel
2. Referent: Privatdozent Dr. Michael Hust
eingereicht am: 05.12.2012
mündliche Prüfung (Disputation) am: 29.04.2013

Druckjahr 2013

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen

Knaust F, Rubelt F, Braig F, Reinhardt R, Konthur Z. A novel universal method for library preparation allowing unidirectional mass-sequencing of amplicons. In Revision.

Rubelt F, Sievert V, Knaust F, Diener C, Lim TS, Skriner K, Klipp E, Reinhardt R, Lehrach H, Konthur Z. Unbiased pyrosequencing of human immunoglobulin mRNA repertoires defines onset of immune senescence. PLoS One. 2012;7(11):e49774. Epub 2012 Nov 30.

Weßling R, Schmidt SM, Micali CO, Knaust F, Reinhardt R, Neumann U, Ver Loren van Themaat E, Panstruga R. Transcriptome analysis of enriched Golovinomyces orontii haustoria by deep 454 pyrosequencing. Fungal Genet Biol. 2012 Jun;49(6):470-82. Epub 2012 Apr 11.

Danksagung

Mein besonderer Dank gilt Herrn Prof. Dr. Dübel für die Möglichkeit zur Promotion, die Förderung und wissenschaftliche Betreuung.

Dr. Richard Reinhardt (Max-Planck-Institut für molekulare Genetik, Berlin und Max-Planck-Institut für Pflanzenzüchtungsforschung, Köln) danke ich für die Hilfestellungen und vielfältigen technischen Möglichkeiten, die mir in seiner Arbeitsgruppe geboten wurden. Bei Dr. Zoltán Konthur (Max-Planck-Institut für molekulare Genetik) bedanke ich mich für die fortwährende Unterstützung meiner Arbeit. Florian Rubelt (Max-Planck-Institut für molekulare Genetik) danke ich für die gute Zusammenarbeit, sowohl im Labor als auch außerhalb. Bei Sven Klages (Max-Planck-Institut für molekulare Genetik) bedanke ich mich für die Unterstützung auf bioinformatischer Seite und die schnelle Hilfe bei der Wiedererlangung älterer Datensätze.

Inhaltsverzeichnis

Inhaltsverzeichnis	I
Abbildungsverzeichnis	V
Tabellenverzeichnis	VIII
Abkürzungsverzeichnis	X
1. Einleitung	1
1.1 Maxam und Gilbert	1
1.2 Sanger	2
1.3 Next Generation Sequencing	5
1.4 Pyrosequencing	6
1.4.1 454-Sequencing	8
1.4.2 Ablauf des 454-Sequenzierens	9
1.4.2.1 emPCR	9
1.4.2.2 Beadrecovery	12
1.4.2.3 Sequencing	13
1.4.3 Data-Processing	15
1.4.3.1 Imageprocessing	15
1.4.3.2 Signalprocessing	15
1.4.4 Library-Arten	17
1.4.4.1 Titanium General Library	18
1.4.4.2 Titanium Paired End Library	20
1.4.4.3 Titanium Rapid Library	21
1.4.4.4 Titanium Amplicon Library	22
1.4.5 Berechnung, cpb-Zahl und Titration	24
1.5 Weitere NGS-Methoden	24
1.5.1 Helicos Bioscience	25
1.5.2 Illumina	25
1.5.3 Life/APG	26
1.5.4 Pacific Bioscience	28
1.5.5 Ion Torrent	29
1.6 Vergleich der Sequencing-Kosten	30
1.7 SPRI	31
1.8 Zielsetzung dieser Arbeit	32
2. Material und Methoden	33
2.1 Material	33
2.1.1 Verbrauchsmaterial	33
2.1.2 Geräte und Laborausstattung	33
2.1.3 Chemikalien und Enzyme	34
2.1.4 Kit-Systeme	34
2.1.5 Oligonukleotide und DNA-Marker	35
2.1.6 Verwendete Puffer und Lösungen	36
2.2 Methoden	37

2.2.1 emPCR und Beadrecovery	37
2.2.1.1 Vorbereitung	37
2.2.1.2 Pre-Emulsion	37
2.2.1.3 Live-Amplification Mix	38
2.2.1.4 Library Capture	38
2.2.1.5 Emulsionsbildung	38
2.2.1.6 Amplifikation	38
2.2.1.7 Emulsions-Kollektion	38
2.2.1.8 Waschen und Beadrecovery	39
2.2.1.9 Enrichment-Vorbereitung	39
2.2.1.10 Vorbereitung der Enrichment-Beads	40
2.2.1.11 Anreicherung der DNA-tragenden Capture-Beads	40
2.2.1.12 Aufreinigung der angereicherten Beads	40
2.2.2 Sequencing	41
2.2.2.1 Vorbereitung	41
2.2.2.2 Pre-Wash	41
2.2.2.3 Vorbereitung der PTP	41
2.2.2.4 Vorbereitung der Packing-Beads	41
2.2.2.5 Vorbereitung der DNA-Beads	42
2.2.2.6 Vorbereitung von Enzyme- und PPIase-Beads	42
2.2.2.7 Zusammenbau und Beladen der PTP	42
2.2.2.8 Vorbereitung des GS Sequencers und Runstart	43
2.2.2.9 Signalprocessing und MID-Analyse	43
2.2.3 SPRI-Aufreinigung	43
2.2.4 Ampurebead-Titration	44
2.2.5 Herstellung von MID-Adapttern	44
2.2.6 Sequenzierungsversuche des aqf-Projektes	45
2.2.6.1 Erstellung von Titanium General Librarys	45
2.2.6.2 Titration der Titanium General Librarys	46
2.2.6.3 Ansatz der emPCR mit den General Librarys	48
2.2.6.4 Laufstart mit den General Librarys	48
2.2.6.5 Erstellung von Amplicon Librarys	49
2.2.6.6 Titration der Amplicon Librarys	51
2.2.6.7 Ansatz der emPCR mit den Amplicon Librarys	51
2.2.6.8 Laufstart mit den Amplicon Librarys	51
2.2.6.9 Amplifikations-Test mit den aqf-Amplicon Librarys	52
2.2.6.10 Amplifikations-Tests mit aqf-Sanger-Klonen	52
2.2.7 Entwicklung einer neuen Methode zur gerichteten Sequenzierung	54
2.2.7.1 Erstellung einer Test-Library	54
2.2.7.2 emPCR mit der Test-Library	55
2.2.7.3 Laufstart mit der Test-Library	55
2.2.7.4 Erstellen von Librarys mit der neuen Methodik aus aqf-cDNA	56
2.2.7.5 emPCR der ersten sechs aqf-Librarys	56

2.2.7.6 Laufstart mit den ersten sechs aqf-Librarys	57
2.2.7.7 emPCR mit weiteren sechs aqf-Librarys	57
2.2.7.8 Laufstart mit den weiteren aqf-Librarys	57
2.2.7.9 Erweiterung auf andere Templates (Sequenzierung einer Illumina-Library)	57
2.2.7.10 Antikörpersequenzierung von schweren Immunglobulinketten	59
2.2.7.11 Verbesserungen des Protokolls für die Antikörpersequenzierung	62
3. Ergebnisse	64
3.1 Ampurebead-Titration	64
3.2 Sequenzierungsversuche des aqf-Projektes	68
3.2.1 Sequenzierung mittels Titanium General Library-Methodik	68
3.2.1.1 Erstellung der Titanium General Librarys	68
3.2.1.2 Titration der General Librarys	71
3.2.1.3 Beadrecovery der General Librarys	73
3.2.1.4 Sequencing-Run der General Librarys	74
3.2.2 Sequenzierung mittels Amplicon-Methodik	78
3.2.2.1 Erstellung der Amplicon Librarys	78
3.2.2.2 Titration der Amplicon Librarys	81
3.2.2.3 Beadrecovery der Amplicon Librarys	82
3.2.2.4 Sequencing Run der Amplicon Librarys	83
3.2.2.5 Amplifikations-Test mit den aqf-Amplicon Librarys	86
3.2.2.6 Amplifikations-Tests mit aqf-Sanger-Klonen	87
3.3 Entwicklung einer neuen Methode zur gerichteten Sequenzierung	91
3.3.1 Erprobung einer Testlibrary	91
3.3.1.1 Erstellung der Testlibrary	91
3.3.1.2 Beadrecovery der emPCR mit der Test-Library	92
3.3.1.3 Sequencing-Run mit der Test-Library	92
3.3.2 Sequenzieren der aqf-Librarys mit der neuen Methodik	95
3.3.2.1 Erstellen von aqf-cDNA-Librarys	95
3.3.2.2 Beadrecovery der ersten sechs aqf-Librarys	96
3.3.2.3 Sequencing-Run mit den ersten sechs aqf-Librarys	97
3.3.2.4 Beadrecovery der weiteren sechs aqf-Librarys	102
3.3.2.5 Sequencing-Run mit den weiteren sechs aqf-Librarys	102
3.3.3 Erweiterung auf andere Templates	107
3.3.3.1 Sequenzierung einer Illumina-Library	107
3.3.3.2 Antikörpersequenzierung	111
3.3.3.3 Verbesserungen des Protokolls für die Antikörpersequenzierung	116
4. Diskussion	121
4.1 Ampurebead-Titration	121
4.2 Sequenzierungsversuche des aqf-Projektes	122
4.2.1 Sequenzierung von aqf-cDNAs über die Erstellung einer General Library	123
4.2.1.1 Erstellung der Titanium General Library	124
4.2.1.2 Titration der General Librarys	125
4.2.1.3 Beadrecovery der General Librarys	127

4.2.1.4 Sequencing Run der General Librarys	128
4.2.2 Sequenzierung von aqf-cDNAs über die Erstellung von Amplicon Librarys	130
4.2.2.1 Erstellung der Amplicon Librarys.....	131
4.2.2.2 Titration der Amplicon Librarys.....	132
4.2.2.3 Beadrecovery der Amplicon Librarys	133
4.2.2.4 Sequencing Run der Amplicon Librarys	133
4.2.2.5 Amplifikations-Test mit den aqf-Amplicon Librarys	135
4.2.2.6 Amplifikations-Tests mit aqf-Sanger-Klonen	136
4.3 Entwicklung einer neuen Methode zur gerichteten Sequenzierung.....	138
4.3.1 Erstellung einer Test-Library	138
4.3.1.1 Test-PCR.....	139
4.3.1.2 emPCR und Beadrecovery der Test-Library	140
4.3.1.3 Sequencing-Run mit der Test-Library	140
4.3.2 Sequenzieren der aqf-Librarys mit der neuen Methodik	141
4.3.2.1 Erstellen der Librarys	141
4.3.2.2 emPCR und Beadrecovery der ersten sechs aqf-Librarys.....	142
4.3.2.3 Sequencing Run mit den ersten sechs aqf-Librarys	142
4.3.2.4 emPCR und Beadrecovery der weiteren sechs aqf-Librarys	144
4.3.2.5 Sequencing Run mit den weiteren sechs aqf-Librarys.....	144
4.3.3 Erweiterung auf andere Templates	146
4.3.3.1 Sequenzierung einer Illumina-Library	146
4.3.3.2 Antikörpersequenzierung	148
4.3.3.3 Verbesserungen des Protokolls für die Antikörpersequenzierung	151
4.4 Ausblick.....	154
5. Zusammenfassung – <i>Summary</i>	156
6. Literatur	158

Abbildungsverzeichnis

Abb. 1: Sequenzierung durch chemische Spaltung	2
Abb. 2: Kettenabbruchmethode nach Sanger	4
Abb. 3: Shotgun-Klonierung	4
Abb. 4: Automatisiertes Hochdurchsatzsystem	5
Abb. 5: Enzymkaskade beim Pyrosequencing	7
Abb. 6: GS FLX Instrument	8
Abb. 7: 454 Library mit Adaptersequenzen	9
Abb. 8: Capture-Beads	9
Abb. 9: Ein effektiver DNA-Strang pro Capture-Bead	10
Abb. 10: Ausbildung der Emulsion	10
Abb. 11: Library-Capture	10
Abb. 12: PCR an Capture-Beads	11
Abb. 13: PCR außerhalb der Capture-Bead	11
Abb. 14: Capture-Bead mit PCR-Produkten	11
Abb. 15: Capture-Bead mit einzelsträngigem Amplifikationsprodukt	12
Abb. 16: Binden an Enrichment-Beads	12
Abb. 17: Waschschritte mit dem MPC	12
Abb. 18: Capture-Beads nach den Waschschritten im MPC	12
Abb. 19: Ende des Beadrecovery	13
Abb. 20: Bead-Layer	14
Abb. 21: Enzyme-, Capture- und Packing-Beads in der PTP (www.roche-applied-science.com)	14
Abb. 22: Vier Bead-Layer (www.roche-applied-science.com)	14
Abb. 23: 2 Regionen	14
Abb. 24: 4 Regionen	14
Abb. 25: Sequencing an den Beads	15
Abb. 26: Ergebnis des Sequencings	15
Abb. 27: Herstellung einer Titanium General Library	19
Abb. 28: Profil einer General Library [85]	20
Abb. 29: Herstellung einer Titanium Paired End Library	20
Abb. 30: Herstellung einer Rapid Library	21
Abb. 31: Prinzip der Herstellung einer Amplicon Library	22
Abb. 32: Prinzip der Sequenzierung mit dem HeliScope System	25
Abb. 33: Prinzip der Solexa-Sequenzierung	26
Abb. 34: Prinzip der SOLID-Sequenzierung	27
Abb. 35: SMRT-Methode von Pacific Bioscience	28
Abb. 36: Bindende Bereiche der Amplicon-Primer	49
Abb. 37: Erstellung von 454-Librarys mittel gerichteter Ligation	54
Abb. 38: Heranamplifizieren der Sfil-Sites	58

Abb. 39: Prinzip der Antikörpersequenzierung	61
Abb. 40: emPCR-Ansatz mit den Librarys hergestellt nach verbessertem Protokoll	63
Abb. 41: Übersicht des Ergebnisteils	64
Abb. 42: Ampurebead-Titration auf dem DNA 7500 Chip	66
Abb. 43: Fortsetzung der Ergebnisse der Ampurebead-Titration auf dem DNA 7500 Chip	67
Abb. 44: Ergebnisse der aqf-cDNAs auf dem 7500 DNA-Chip	68
Abb. 45: aqf-Titanium General Librarys auf dem RNA Pico 6000 Chip	70
Abb. 46: Graphische Darstellung der Titrationsergebnisse.....	73
Abb. 47: Größenverteilung der Reads beim Run mit den aqf-General-Librarys	76
Abb. 48: Größenverteilung der Kontroll-DNA beim Run mit den aqf-General-Librarys	76
Abb. 49: Prozentuale Verteilung beim Run mit den aqf-General-Librarys	77
Abb. 50: Graphische Darstellung der Kontroll-DNA-Analyse	77
Abb. 51: Gelbild der Gradienten-PCR von aqf3b und aqf 4b	78
Abb. 52: Gelbild der Gradienten-PCR von aqf5b und aqf6b	79
Abb. 53: Gelbild der Gradienten-PCR von aqf9b und aqf10b	79
Abb. 54: Graphische Darstellung der Amplicon Library-Titrationsergebnisse	82
Abb. 55: Größenverteilung der Reads beim Run mit den aqf-Amplicon Librarys	85
Abb. 56: Größenverteilung der Kontroll-DNA beim Run mit den aqf-Amplicon Librarys	85
Abb. 57: Prozentuale Verteilung beim Run mit den aqf-Amplicon Librarys	85
Abb. 58: Graphische Darstellung Kontroll-DNA-Analyse	86
Abb. 59: Gelbild des Amplifikations-Tests.....	86
Abb. 60: Erstellung von Amplicon Librarys mit aqf-Sanger-Klonen	87
Abb. 61: Amplifikation der Gradienten-PCR-Produkte mit 454-Primern	88
Abb. 62: Two-Step-PCR -erste Amplifikation	89
Abb. 63: Two-Step-PCR –zweite Amplifikation	89
Abb. 64: Gelbild der PCR mit der Testlibrary	91
Abb. 65: Größenverteilung der Reads beim Run mit der Test-Library.....	93
Abb. 66: Prozentuale Verteilung beim Run mit der Test-Library.....	94
Abb. 67: Graphische Darstellung der Kontroll-DNA-Analyse	94
Abb. 68: Längenverteilung der Test-Library	95
Abb. 69: Gelbild der Test-PCR mit den aqf-454-Librarys.....	96
Abb. 70: Größenverteilung der Reads beim Run der ersten sechs aqf-Librarys	98
Abb. 71: Größenverteilung der Kontroll-DNA-Reads beim Run der ersten sechs aqf-Librarys	99
Abb. 72: Prozentuale Verteilung beim Run der ersten sechs aqf-Librarys	99
Abb. 73: Graphische Darstellung der Kontroll-DNA-Analyse	100
Abb. 74: Größenverteilungen von aqf3b und aqf4b	100
Abb. 75: Größenverteilungen von aqf5b, aqf6b, aqf9b und aqf10b	101
Abb. 76: Größenverteilung der Reads beim Run mit den weiteren sechs aqf-Librarys	104
Abb. 77: Größenverteilung der Kontroll-DNA-Reads beim Run mit den weiteren aqf-Librarys	104
Abb. 78: Prozentuale Verteilung beim Run mit den weiteren aqf-Librarys	104
Abb. 79: Graphische Darstellung der Kontroll-DNA-Analyse	105

Abb. 80: Größenverteilungen von aqf15 und aqf25	105
Abb. 81: Größenverteilungen von aqf35, aqf45, aqf55, aqf65	106
Abb. 82: Größenverteilung der Reads beim Run mit der OCon-Library	109
Abb. 83: Prozentuale Verteilung beim Run mit der OCon-Library	109
Abb. 84: Graphische Darstellung der Kontroll-DNA-Analyse	110
Abb. 85: Längenverteilungen der OCon-Library.....	110
Abb. 86: Größenverteilung der Antikörper-Librarys	111
Abb. 87: Größenverteilung der Antikörper-Library-Reads.....	113
Abb. 88: Größenverteilungen der Kontroll-DNA-Reads	113
Abb. 89: Prozentuale Verteilung beim Run mit den Antikörper-Librarys.....	114
Abb. 90: Graphische Darstellung der Kontroll-DNA-Analyse	114
Abb. 91: Größenverteilungen der Antikörper-Librarys im Run	115
Abb. 92: Konzentrationsprofil der Library RA-BS_10LK	117
Abb. 93: Konzentrationsprofil der Library KaretiLK	117
Abb. 94: Konzentrationsprofil der Library KaretiLK nach optionalem Größenunterschluss	117
Abb. 95: Größenverteilung der Libraryreads	118
Abb. 96: Größenverteilung der Kontroll-DNA-Reads	119
Abb. 97: Prozentuale Verteilung beim Run	119
Abb. 98: Graphische Darstellung der Kontroll-DNA-Analyse	120

Tabellenverzeichnis

Tab. 1: Ausbeuten pro Region [75]	13
Tab. 2: Ausbeuten von Amplicon Librarys [94].....	23
Tab. 3: Sequenzierstatistiken von humanen Genomprojekten [111]	30
Tab. 4: Sequenzierkosten pro Gb (Kommunikation mit Richard Reinhardt)	31
Tab. 5: Berechnung der Titrationsansätze der General Librarys	47
Tab. 6: emPCR-Ansatz der vier aqf General Librarys.....	48
Tab. 7: Zusammensetzung der PTP	49
Tab. 8: Primerkombinationen bei der Amplicon Library-Herstellung.....	50
Tab. 9: Zusammensetzung eines PCR-Ansatzes	50
Tab. 10: Berechnung der Titration der Amplicon Librarys.....	51
Tab. 11: emPCR-Ansatz der aqf-Amplicon Librarys	51
Tab. 12: Beladung beider PTP-Regionen des Amplicon Laufs.....	52
Tab. 13: PCR-Ansatz des Amplifikations-Tests	52
Tab. 14: emPCR-Ansatz mit der Test-Library	55
Tab. 15: Verwendete Adaptersequenzen.....	56
Tab. 16: emPCR-Ansatz von aqf3b bis aqf10b	56
Tab. 17: emPCR-Ansatz von aqf15 bis aqf65	57
Tab. 18: PCR-Ansatz zum Heranamplifizieren der Sfil-Sites.....	58
Tab. 19: Berechnung der Peak-Ratios bei der Ampurebead-Titration	65
Tab. 20: Titration der General Librarys	72
Tab. 21: Beadrecovery der emPCR von den General Librarys.....	73
Tab. 22: Run mit den aqf-General Library.....	74
Tab. 23: Kontroll-DNA-Analyse beim Run mit den aqf-General-Librarys.....	77
Tab. 24: Titration der Amplicon Librarys	81
Tab. 25: Beadrecovery der Amplicon Librarys	83
Tab. 26: Run mit den aqf-Amplicon Librarys	84
Tab. 27: Kontroll-DNA-Analyse beim Run mit den aqf-Amplicon Librarys.....	85
Tab. 28: Beadrecovery der emPCR mit der Test-Library	92
Tab. 29: Run mit der Test-Library	93
Tab. 30: Kontroll-DNA-Analyse beim Run mit der Test-Library	94
Tab. 31: Zerlegung in die MIDs beim Run mit der Test-Library von Region 1	95
Tab. 32: Konzentrationsbestimmung der aqf-Librarys	96
Tab. 33: Beadrecovery der Librarys aqf3b bis aqf10b	96
Tab. 34: Run der ersten sechs aqf-Librarys	98
Tab. 35: Kontroll-DNA-Analyse beim Run der ersten sechs aqf-Librarys.....	99
Tab. 36: Zerlegung in die MIDs beim Run der ersten sechs aqf-Librarys.....	100
Tab. 37: Beadrecovery der Librarys aqf15 bis aqf65	102
Tab. 38: Run mit den weiteren sechs aqf-Librarys.....	103

Tab. 39: Kontroll-DNA-Analyse beim Run mit den weiteren aqf-Librarys	104
Tab. 40: Zerlegung in die MIDs beim Run der weiteren aqf-Librarys	105
Tab. 41: Runs mit der OCon-Library	108
Tab. 42: Kontroll-DNA-Analyse beim Run mit der OCon-Library	109
Tab. 43: Run der Antikörpersequenzierung	113
Tab. 44: Kontroll-DNA-Analyse des Runs mit den Antikörper-Librarys	114
Tab. 45: Zerlegung in die MIDs beim Run mit den Antikörper-Librarys	114
Tab. 46: Mengenvergleich vor und nach Libraryproduktion	116
Tab. 47: Run mit Librarys nach verbessertem Protokoll	118
Tab. 48: Kontroll-DNA-Analyse	119

Abkürzungsverzeichnis

A	Adenin
APS	Adenosin-5'-Phosphosulfat
ATP	Adenosin-Tri-Phosphat
BAC	Bacterial Artificial Chromosome
BB2	Bead Buffer 2
BDD	Bead Deposition Device
BCCP	Biotin-Carboxyl-Carrier-Protein
bp	Basenpaare
C	Cytosin
CCD	Charge Coupled Device
cDNA	komplementäre DNA (englisch <i>complementary DNA</i>)
contig	Contiguous Sequences
cpb	Copy Per Bead
CRT	Cyclic Reversible Termination
dATP	Desoxyadenosin-Triphosphat
DBIM	DNA Bead Incubation Mix
dCTP	Desoxycytidin-Triphosphat
ddaH ₂ O	doppelt destilliertes und autoklaviertes Wasser
ddNTP	Dideoxynukleosid-Triphosphat
dGTP	Desoxyguanosin-Triphosphat
DMS	Dimethylsulfat
DNA	Desoxyribonukleinsäure (englisch <i>deoxyribonucleic acid</i>)
dNTP	Deoxynukleosid-Triphosphat
dTTP	Desoxythymidin-Triphosphat
emPCR	Emulsions-PCR
EST	Expressed Sequence Tag
FU	Fluoreszenz-Unit
G	Guanin
g	Gramm
<i>g</i>	Erdbeschleunigung (9,81 m/s ²)
Gb	Gigabasen
h	Stunden
Hz	Hertz
I	Inosin
IgH	schwere Immunglobulinkette
kb	Kilobasen
l	Liter
m	Meter

M	Molar (mol/l)
mAU	Milli-Absorbance-Unit
Mb	Megabasen
MID	Multiplex Identifier
min	Minuten
Mio.	Million
MPC	Magnetic Particle Collector
MPI	Max-Planck-Institut
NaCl	Natriumchlorid
NGS	Next Generation Sequencing
PCR	Polymerase Chain Reaction
PPi	Pyrophosphat
PTP	Picotiterplatte
RNA	Ribonukleinsäure
s	Sekunden
SMRT	Single Molecule Realtime
SNP	Single Nucleotide Polymorphism
SPRI	Solid Phase Reversible Immobilization
SSB	Single Stranded DNA Binding Protein
sscDNA	einzelsträngige cDNA
T	Thymin
t	Zeit
U	Unit
ZMV	Zero-Mode Waveguide Detector

1. Einleitung

Im Jahr 1869 isolierte Friedrich Mistier aus Zellkernen (Nuclei) von Leukozyten eine Substanz, die er daraufhin Nukleinsäuren nannte [1]. Erst 1944 konnten Oswald Avery, Colin MacLeod und Maclyn McCarty nach zehnjährigen Untersuchungen zeigen, dass Desoxyribonukleinsäure (DNA) der Träger genetischer Informationen ist [2]. Die chemische Struktur der DNA wurde durch die Arbeiten von Phoebus Levine und Alexander Todd aufgeklärt [3,4], und von Erwin Chargaff um die „Chargaff-Regel“, der fixen Relation A gleich T und G gleich C ergänzt [5]. James Watson und Francis Crick klärten die dreidimensionale Struktur der DNA 1953 auf, indem sie die damaligen Forschungsergebnisse verknüpften, und konstruierten das Modell der Doppelhelix der DNA [6], welches durch Röntgenstrukturanalysen von Rosalind Franklin [7] und Maurice Wilkins bestätigt wurde [8]. Crick war es auch, der 1958 die Beziehungen zwischen DNA, RNA und Proteinen als das zentrale Dogma der Molekularbiologie zusammenfasste [9]; der genetische Code wurde dann 1961 von Marshall Nirenberg entschlüsselt [10].

Diese Arbeiten stellen die grundlegende Voraussetzung für die Entwicklung von Technologien dar, um die DNA *in vitro* sequenzieren zu können. Im Lauf der letzten 35 Jahre gab es beständigen Fortschritt und Erfolge, die Basenabfolge der DNA zu bestimmen.

1.1 Maxam und Gilbert

Walter Gilbert und Allan Maxam veröffentlichten 1977 eine Methode zur Sequenzierung von DNA, die auf chemischer Spaltung beruht [11]. Dabei wird die zuvor radioaktiv markierte DNA in verschiedenen Ansätzen unterschiedlichen Chemikalien und Versuchsbedingungen ausgesetzt, was zu basenspezifischen Spaltungen führt. Die Konzentrationen sind so gewählt, dass es durchschnittlich nur zu einer Spaltung pro DNA-Molekül kommt. Die in jedem Ansatz entstandenen Fragmente werden dann auf ein denaturierendes Polyacrylamid-Gel aufgetragen, über einen Röntgenfilm sichtbar gemacht und so wird die Basenabfolge rekonstruiert (Abb. 1 auf Seite 2).

Maxam und Gilbert konnten so ca. 100 bp lange Fragmente sequenzieren. Das Sequenzieren durch chemische Spaltung verlor im Lauf der Zeit zunehmenden an Interesse aufgrund des Umgangs mit den teilweise hochgiftigen Chemikalien und dem Fortschritt anderer Sequenziermethoden.

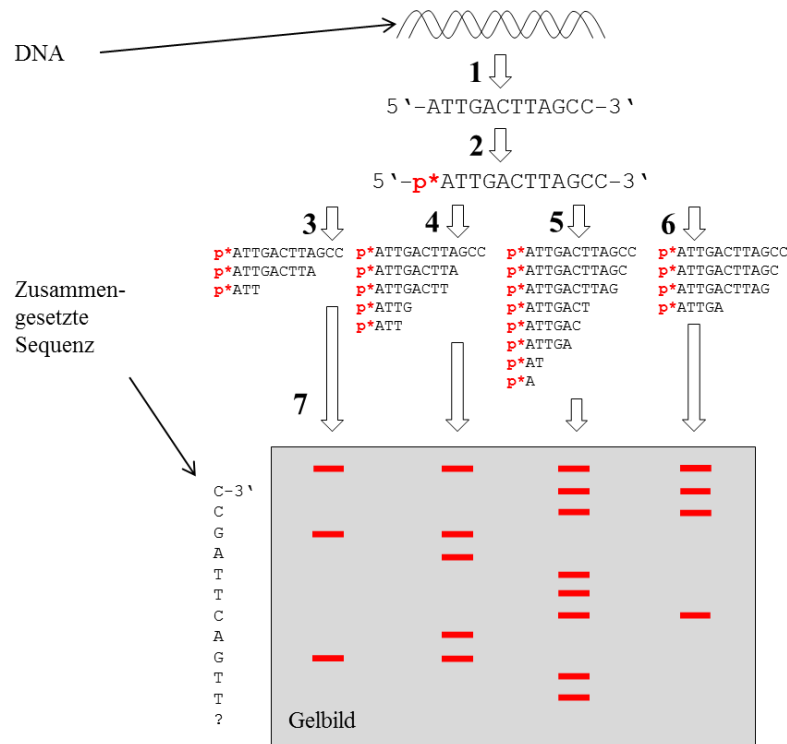


Abb. 1: Sequenzierung durch chemische Spaltung

Hierbei wird die DNA zerkleinert (beispielsweise mit Restriktionsenzymen) und denaturiert, um Einzelstränge zu erhalten (1). Um diese radioaktiv zu markieren, werden die terminalen Phosphate mit alkalischer Phosphatase entfernt, und das 5'-Ende mit $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ durch eine Polynukleotidkinase markiert (2). Bei der Behandlung eines Ansatzes mit DMS zur Methylierung unter alkalischen Bedingungen und Piperidin wird die glycosidische Bindung von Guanin gespalten (3). Bei Einsatz von DMS im Sauren und Piperidin werden die glycosidischen Bindungen von Guanin und Adenin gespalten (4). Hydrazin und Piperidin führen zur Spaltung der glycosidischen Bindungen vor Thymin und Cytosin (5); wird die DNA mit Hydrazin in einer NaCl-Lösung behandelt, so wird die Reaktion von Thymin mit Hydrazin unterdrückt, es findet nur eine glycosidische Spaltung von Cytosin statt (6). Nach der parallelen elektrophoretischen Auftrennung der Ansätze wird ein Autoradiogramm angefertigt, auf dem die einzelnen Banden sichtbar werden (7). Aufgrund der Laufweiten der einzelnen Banden kann die Sequenz (links neben dem Gelbild) rekonstruiert werden [12].

1.2 Sanger

Ebenfalls im Jahr 1977 veröffentlichte Frederick Sanger die Kettenabbruchmethode [13], die auf der Verwendung von Dideoxy-Nukleosid-Triphosphaten (ddNTPs) basiert, denen die 3'-OH-Gruppe fehlt; wird ein ddNTP anstelle eines dNTPs von einer DNA-Polymerase in den entstehenden DNA-Strang eingebaut, so kann keine weitere Verlängerung des DNA-Strangs erfolgen. Es kommt zum Kettenabbruch. Die Kettenabbruchmethode stellt eine Weiterentwicklung der von Sanger 1975 veröffentlichten plus/minus-Sequenzieretechnik dar [14], in der für die Kettenabbruchmethode folgende wichtige Bestandteile beschrieben wurden: die Verwendung von Primern als Startpunkt für eine DNA-Polymerase, basenspezifischer Kettenabbruch und denaturierende Polyacrylamidgele, die die Unterscheidung von DNA-Fragmenten um nur ein Basenpaar durch elektrophoretische Auftrennung ermöglichen.

Bei der ursprünglichen Kettenabbruchmethode musste die DNA auf vier Ansätze verteilt werden, die alle die einzelsträngige DNA als Matrize, Primer als Ansatzpunkt für die DNA-Polymerase, die vier dNTPs (dATP, dCTP, dGTP und dTTP, wobei mindestens eins radioaktiv markiert ist [α - ^{32}P]) und je eins der vier ddNTPs enthalten. Die Konzentration der ddNTPs wird dabei so gering eingestellt, dass es statistisch nur zu einem Einbau pro Strangsynthese auf der Matrize kommt (und damit zum Abbruch). Nach Auftrennung der vier Ansätze auf einem denaturierenden Polyacrylamid-Gel und Anfertigen eines Autoradiogramms können so wie bei der chemischen Spaltung die Basenpositionen bestimmt werden.

Die Entwicklung der PCR [15] führte zu einer Verbesserung der Sanger-Sequenzierung, da jetzt analog der PCR Sequenzierreaktionen in Thermocyclern als Cycle-Sequencing durchgeführt werden konnten. Durch die Entwicklung von Fluoreszenz-Farbstoffen [16,17] konnte auf den Einsatz radioaktiv markierter dNTPs verzichtet werden. Da die 5'-Enden von Primern mit Fluoreszenz-Farbstoffen markiert wurden, nennt man diese Methode „Dye-Primer-Sequencing“. Durch die Markierung der ddNTPs selber mit unterschiedlichen Fluoreszenz-Farbstoffen [18], genannt „Dye-Terminator“, kann die Sequenzierung eines DNA-Stücks in nur einem Ansatz durchgeführt werden. Nach der Entwicklung der Kapillargelelektrophorese [19,20] Anfang der 90er Jahre für die Auftrennung von Sequenzierprodukten nach Sanger's Kettenabbruchmethode kamen bald entsprechende halbautomatische Kapillar-Sequenziergeräte auf, mit denen fast ausschließlich die gesamte DNA-Sequenzierung durchgeführt wurden (Abb. 2 auf Seite 4).

Mit der Kettenabbruchmethode nach Sanger lassen sich Leseweiten von bis zu 1000 bp erreichen; für größere DNA-Moleküle mussten Klonierungsstrategien entwickelt werden, deren bekanntester Vertreter die Shotgun-Klonierung ist [21]. Die DNA wird hierbei enzymatisch oder mechanisch in kleinere Stücke zerteilt (Abb. 3 auf Seite 4) und es entstehen zufällig überlappende Fragmente, aus denen nach deren Sequenzierung das ursprüngliche große DNA-Molekül wieder zusammengebaut werden kann.

Ein limitierender Faktor beim Probendurchsatz der Sangersequenzierung ist die Herstellung der zu sequenzierenden DNA (Template) in ausreichender Menge und Qualität. Eine Möglichkeit stellt die PCR dar. Es wurden auch am MPI für molekulare Genetik von Holger Rauth und Richard Reinhardt verschiedene Verfahren entwickelt für die automatisierte Template-Präparation [22]. Ein Problem dieser Variante ist allerdings der sogenannte Amplifikation-Bias; bestimmte DNA-Sequenzen werden mittels PCR nur schlecht oder gar nicht amplifiziert. Eine Alternative hierzu ist die Generierung von Template-DNA *in vivo* aus einer Klon-Bibliothek. Die in Plasmiden ligierte Template-DNA wird durch Zellkultivierung amplifiziert, wobei die Größe der Konstrukte durch Vektortyp und Replikationseffizienz bestimmt werden. Somit wird ein Amplifikations-Bias vermieden und darüber hinaus ist die Qualität von aufgereinigter Template-DNA wesentlich höher. Hierfür und für die Automatisierung weiterer Arbeitsschritte wurde ebenfalls im MPI für molekulare Genetik von Heiner Kuhl [23] ein automatisiertes Hochdurchsatzsystem entwickelt (Abb. 4 auf Seite 5).

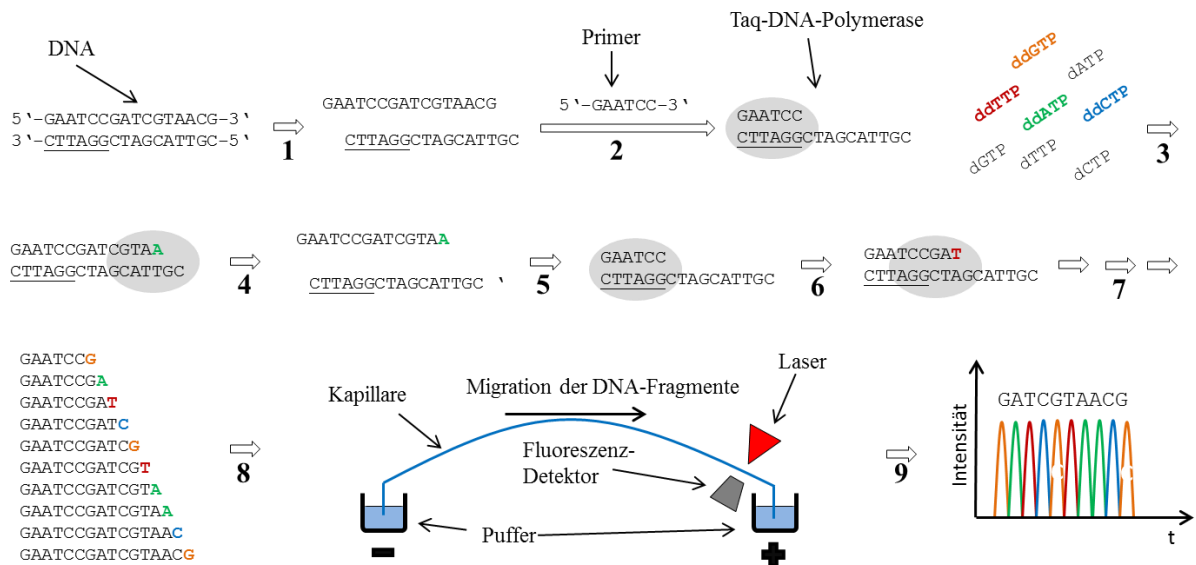


Abb. 2: Kettenabbruchmethode nach Sanger

Die doppelsträngige DNA wird durch Hitze denaturiert (1). Die Sequenz eines Teilbereichs der DNA muss bekannt sein (unterstrichen), damit sich daran ein kurzes, komplementäres DNA-Stück, genannt Primer, anlagern kann (2); dies ist unerlässlich für die Bindung von jeglichen DNA-Polymerasen. Die Anlagerung des Primers an die DNA-Matrize nennt man Annealing. Die hierbei verwendeten Polymerasen müssen hitzestabil sein (z.B. Taq-Polymerase), damit entsprechend der PCR ein Cycle-Sequencing stattfinden kann. Im Sequenziermix sind dNTPs und die mit unterschiedlichen Fluoreszenz-Farbstoffen markierten ddNTPs vorhanden; die Polymerase beginnt nun in der Elongationsphase anhand der Matrize den Gegenstrang aufzubauen (3), bis zufällig anstelle eines dNTPs das entsprechende ddNTP eingebaut wird. Dadurch stoppt die Elongation und es kommt zum Kettenabbruch. Es folgen nun eine bestimmte Anzahl Wiederholungen von Denaturierung (4), Annealing (5) und Elongation (6), so dass ein Gemisch von unterschiedlich langen Fragmenten entsteht (7), die anhand der gleichen DNA-Matrize hergestellt wurden. Dieses Fragment-Gemisch wird aufgereinigt und per Kapillarelektrophorese aufgetrennt (8); man erhält ein Elektropherogramm (9), aus dem die Sequenz errechnet wird.

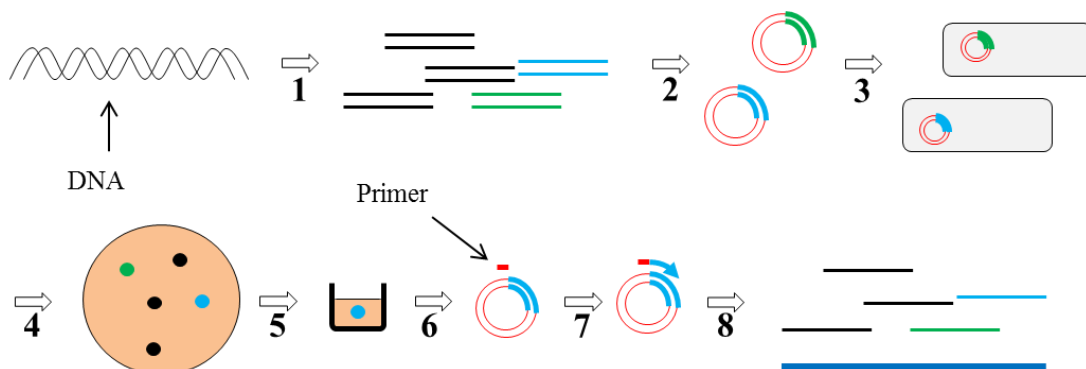


Abb. 3: Shotgun-Klonierung

Die DNA wird in kleinere Stücke zerteilt (1), beispielsweise durch Ultraschall. Die Fragmente (Inserts) werden anschließend in Plasmid-Vektoren, z.B. pUC19, ligiert (2) und mittels Transformation in spezielle *E. coli*-Zellen eingeführt (3), wobei die Bedingungen so gewählt sind, dass prinzipiell nur ein Vektor-DNA-Konstrukt pro Zelle vorhanden ist. Die transformierten Zellen wachsen auf speziellen Nährböden zu Kolonien voran (4); dadurch findet eine Vereinzelung statt, die Kolonien sind monoklonal. Die Klone werden gepickt und vermehrt (5). Nach Isolation der DNA findet über spezielle Primer, die komplementär zu den bekannten Vektorsequenzen sind (6), eine Sanger-Sequenzierung auf den Plasmid-Vektoren der Inserts statt (7). So erhält man eine Vielzahl überlappender Sequenzen, die zu einer Konsensus Sequenz (dunkelblau markiert) zusammengesetzt werden können (8).



Abb. 4: Automatisiertes Hochdurchsatzsystem

Abbildung 4 zeigt das im MPI für molekulare Genetik entwickelte Hochdurchsatzsystem für die Probenvorbereitung zum Sanger-Sequencing mit fahrbarem Roboterarm, automatischen Pipettierstationen, Karussell, Zentrifugen und Kühlschrank. Das System steht seit 2011 im Max-Planck-Genomzenter Köln.

Mit der Kettenabbruchmethode nach Sanger konnte so auch erstmals im Jahr 1995 das komplette Genom eines Lebewesens, das von *Haemophilus influenzae*, bestimmt werden [24]. Mittels der in diesem Kapitel beschriebenen stetigen Verbesserungen der Kettenabbruchmethode war es schließlich möglich, das menschliche Genom von einem internationalen Konsortium nach wenigen Jahren Sequenzierung zu entschlüsseln [25,26].

1.3 Next Generation Sequencing

Das Sequenzieren nach Sanger hat die letzten drei Jahrzehnte dominiert; mit den im vorangegangenen Kapitel beschriebenen Verbesserungen –dem Einsatz von Dye-Terminator-Chemie, PCR und Kapillargelelektrophorese– hatte die Kettenabbruchmethode ihren Zenit erreicht. In den letzten Jahren sind Methoden zur DNA-Sequenzierung aufgekommen, mit denen Sequenzen bis zu 100-mal schneller und wesentlich preiswerter generiert werden können. Diese Methoden werden unter dem Sammelbegriff des „Next Generation Sequencing“ (NGS) zusammengefasst.

Die Herangehensweise bei den unterschiedlichen NGS-Methoden ist teilweise sehr verschieden. Jedoch gibt es aber zwei Gemeinsamkeiten, durch die sie sich vom Sanger-Sequencing unterscheiden. Beim NGS wird keine Klonierung *in vivo* durchgeführt. Dieses ist beim Sanger-Sequenzieren unerlässlich für die Vereinzelung der Klone (vergleiche Abb. 3) und stellt einen materiellen und zeitlichen Engpass dar. Bei den unterschiedlichen NGS-Methoden wird diese Klonierung bzw. Vereinzelung *in vitro* durchgeführt oder erst gar nicht benötigt (single molecular sequencing). Außerdem wird beim NGS im Gegensatz zum Sanger-Sequencing das Sequenzieren selber, also der Vorgang des Nukleotideinbaus analysiert und ausgewertet.

In den folgenden Abschnitten werden die derzeit verbreitetsten Verfahren vorgestellt, wobei der Schwerpunkt auf der in dieser Arbeit verwendeten Methode, dem 454-Pyrosequencing gelegt wird.

1.4 Pyrosequencing

Pyrosequencing beruht auf der Freisetzung von Pyrophosphat (PPi) während der DNA-Strangsynthese. Durch eine Abfolge enzymatischer Reaktionen wird Licht erzeugt, dessen Intensität proportional zu der Menge der eingebauten Nukleotide ist. Da das Bestimmen der Basen während des DNA-Strangaufbaus stattfindet, ist das Pyrosequencing eine „Sequencing By Synthesis“-Methode.

Man kann die Prozesse beim Pyrosequencing in mehrere Abschnitte unterteilen. Als erstes bindet ein Primer an die zu sequenzierende einzelsträngige Template-DNA. Dann wird eins der vier dNTPs dazugegeben, welches von der DNA-Polymerase eingebaut wird. Wenn ein dNTP erfolgreich eingebaut wird entsteht Pyrophosphat, wobei die Menge der freigesetzten Pyrophosphate äquimolar der Menge der eingebauten dNTPs ist. Eine ATP-Sulfurylase wandelt PPi mit Adenosin-5'-Phosphosulfat (APS) in ATP um. In der Gegenwart von ATP wird Luciferin durch das Enzym Luciferase in Oxyluciferin umgewandelt, wobei Licht entsteht. Dessen Intensität ist proportional zur ATP-Menge und damit zu freigesetztem PPi bzw. eingebauten dNTPs. Das entstehende Licht wird von einer CCD-Kamera aufgezeichnet und in einem Pyrogramm als Peak dargestellt [27].

Schon 1985 wurden die Prinzipien des „sequencing-by-synthesis“ beschrieben [28], allerdings basierte diese Methode auf der Messung der dNTP- Absorption während des Strangaufbaus, was hohe DNA-Template- und dNTP-Konzentrationen benötigt. Pal Nyrén entwickelte ebenfalls 1985 eine sehr sensitive luminometrische Methode [29], die es ermöglichte, der Pyrophosphat-Synthese in Echtzeit zu folgen [30,31]. 1987 wurde dann erstmals von Nyrén die grundlegende Enzymkaskade entwickelt, bestehend aus DNA-Polymerase, ATP-Sulfurylase und Luciferase, mit der PPi kontinuierlich gemessen werden konnte [32]. Ein Problem dieser Methode war die hohe Hintergrundstrahlung während der Lumineszenz-Messung, wobei die verwendeten Enzyme und dNTPs die Hauptursache waren. Durch die Verwendung eines weiteren Enzyms, einer Apyrase, die unverbrauchte dNTPs und ATP in die entsprechenden Monophosphate umwandelt, konnte die Hintergrundstrahlung verringert werden und es wurden Waschschrte mit auf magnetischen Beads immobilisierte Apyrase eingeführt [33]. Außerdem wurden dNTPs und die Polymerase mit Pyrophosphatasen behandelt, um die in niedrigen Konzentrationen schon vorhandenen PPis abzubauen, was ebenfalls zu einer Verringerung der Hintergrundstrahlung führte. Vorangegangene Entwicklungen von Festphasen-DNA-Sequenzierung [34] -oder „solid-phase-sequencing“- machten es möglich, die Pyrosequencing-Methode auf Festphasen-Systeme zu adaptieren [35]. Ein weiteres Problem bestand darin, dass dATP ebenfalls zu einem geringen Teil ein Substrat für Luciferase ist, was dazu führt, dass auch bei nichterfolgtem Einbau von dATP es trotzdem zu einer falschpositiven Reaktion kommt. Die Lösung dieses Problems bestand in der Verwendung von α -S-modifiziertem Nukleotid -dATP α S, welches wie dATP von DNA-Polymerasen beim Strangaufbau verwendet wird, aber ein nur sehr schlechtes Substrat für Luciferase ist. Mit diesen Modifikationen konnten 1996 erstmals 15 bp sequenziert werden [36]. Eine weitere Verbesserung war die Entwicklung und Verwendung von rekombinanter ATP-Sulfurylase [37], produziert in *E.coli*, womit falsche Signale, deren Ursache in der früheren Sulfurylase-Präparation zu finden waren, beseitigt werden konnten [38].

Nach Anpassen der Reaktionskinetiken und einzelner Optimierungen konnte 1998 ein Protokoll zur Pyrosequenzierung veröffentlicht werden (Abb. 5), das seit dem kaum noch verändert wurde [27,39].

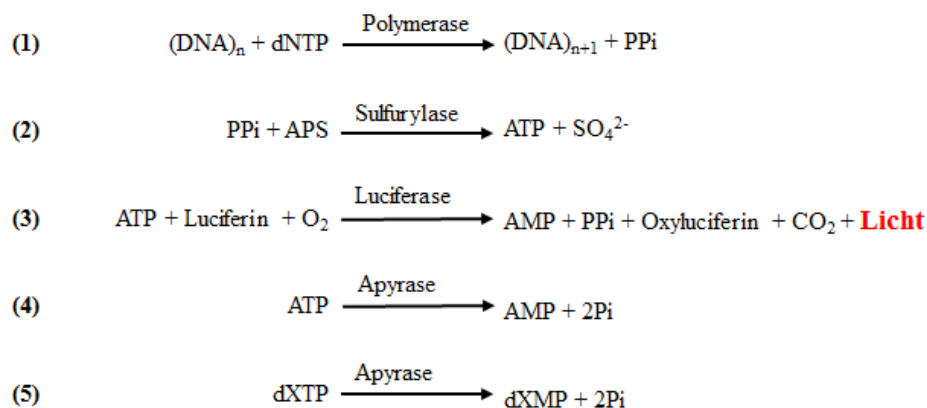


Abb. 5: Enzymkaskade beim Pyrosequencing

Zu der Template-DNA, an der zuvor ein Primer gebunden hat, wird nun ein dNTP gegeben, beispielsweise ein dCTP. Kann dies von einer Polymerase eingebaut werden, weil auf der Template-DNA an dieser Stelle ein G vorhanden ist, so wird ein PPi frei (1). PPi wird mit APS durch eine ATP-Sulfurylase in ATP umgewandelt (2). Durch das entstandene ATP kann die Luciferase jetzt Luciferin in Oxyluciferin umwandeln, wobei Licht frei wird (3). Die Nukleotide werden kontinuierlich von Apyrase degradiert (4+5), um die Hintergrundstrahlung zu vermindern. Diese Abfolge von Reaktionen wird nun mit den anderen dNTPs wiederholt.

Beim Pyrosequenzieren trat das Problem auf, dass manche Template-DNAs sich besser sequenzieren ließen als andere. Als eine mögliche Ursache wurde das Ausbilden von Sekundärstrukturen der DNA angenommen. Um dem entgegenzuwirken, setzte man mit Erfolg Single-Stranded DNA-Binding-Protein (SSB) ein, womit die Qualität des Sequencing verbessert wurde [40,41].

Das erste kommerziell erhältliche automatisierte System wurde 1999 von der Firma Pyrosequencing AB verkauft, basierend auf einer Entwicklung von Nyérén, Nordström und Ronaghi [42]. Es konnten damit 96 Proben parallel sequenziert werden. Dieses System wurde hauptsächlich zur Analyse von Single Nucleotide Polymorphismen (SNPs) verwendet [43]. Später fand es Verwendung bei einer Vielzahl von Anwendungen: der Analyse von Allel-Häufigkeiten [44], Methylierungsanalysen [45], Sequenzierung heteroplasmatischer DNA [46], Typisierung verschiedener Viren, Bakterien und Pilze [47-49] und forensischen Analysen [50]. Der Vorteil vom Pyrosequencing gegenüber Sanger-Sequencing bestand derzeit in den geringeren Kosten von ca. 20 Cent pro Probe [51].

Eine Vergrößerung der Leseweiten konnte dadurch erreicht werden, dass anstelle einer dATP α S-Lösung, in der sich auch das R-Isomer befand, eine Lösung verwendet wurde, in der ausschließlich das S-Isomer enthalten ist, womit eine Hemmung der Apyrase-Aktivität verringert werden konnte; mit dem neuen Ansatz wurden Leseweiten von bis zu 153 bp erreicht [52]. Weitere Verbesserungen folgten; dATP α S wurde durch 7-deaza-dATP ausgetauscht, das nicht die Apyrase-Aktivität verringerte [53]; die bis dahin benutzte Klenow-Polymerase wurde durch Sequenase, eine gentechnisch veränderte T7 Polymerase [54], ersetzt, die besser für die Sequenzierung von homopolymeren

Bereichen geeignet ist [55]; durch die Verwendung von Betain konnte die Temperatur der Sequenzierreaktion auf 37°C gesteigert werden, was eine Erhöhung der Enzymaktivitäten und damit verbunden bessere Leseweiten bedeutete [56,57]. Das Reannealen von übriggebliebenen Primern, was bis dahin ein weiteres Problem beim Pyrosequencing war, konnte durch die Verwendung bestimmter, teilweise zu den Primern komplementären Oligonukleotide verhindert werden; damit einhergehend war es möglich, doppelsträngige DNA in einem Schritt als Template zu präparieren [58,59]. Eine weitere Verbesserung stellte die Verwendung eines Multiprimer-Ansatzes dar [60,47], mit dessen Hilfe u.a. Proben mit unterschiedlichen Genotypen sequenziert werden konnten [61].

1.4.1 454-Sequencing

Die Biotechnologie-Firma 454 Life Sciences entwickelte das erste auf dem Markt erhältliche NGS-Gerät (Abb. 6). Ausgehend von den im vorigen Abschnitt beschriebenen Entwicklungen, begann 454 mit der Miniaturisierung des Pyrosequencing, um den Probendurchsatz zu maximieren.

Um die Sequenzierreaktionen, die auf Solid-Phase-Technologie basieren, zu separieren, entwickelte 454 eine Platte bestehend aus Millionen winziger Reaktionsräume (Wells), genannt Picotiterplatte (PTP) [62]. Die 75 mm x 75 mm großen PTPs besitzen 55 µm tiefe Wells mit einem Durchmesser von 44 µm. Die Herstellung dieser PTPs beruht auf Ätztechniken an (Glas-) Faserkabeln [63-65] und die ersten so hergestellten PTPs besaßen ca. 1,6 Mio. Wells [66].

Mit einem Sequenzier-Lauf (Run) können so ca. 400000 Sequenzen (Reads) mit einer durchschnittlichen Länge von etwas über 200 bp innerhalb 7,5 Stunden generiert werden (altes „Standard FLX“-System). Zur Jahreswende 2008/2009 kam eine weiterentwickelte PTP auf den Markt. Die neue Version, genannt Titanium-PTP, besitzt 3,4 Mio. Wells mit einem Durchmesser von 29 µm und einer Tiefe von 34 µm (Titanium-System). Mit dem Titanium-Format sollen so zwischen 800000-1,2 Mio. Reads mit einer durchschnittlichen Länge von 400 bp pro Run generiert werden können. Neu ist das Titanium plus-Format, das ebenfalls die Titanium-PTP verwendet, mit dem modale Leseweiten bis zu 700 bp erreicht werden sollen.

Um große Leseweiten im PTP-Format zu erreichen, sind zwei Dinge ausschlaggebend: die Substrate für die Enzymkaskade müssen problemlos angeliefert und Reaktionsprodukte dementsprechend auch wieder entfernt werden können; außerdem müssen die verwendeten Enzyme diese Transportprozesse überstehen und dürfen nicht aus den Wells herausgewaschen werden. Letzteres erreichte 454 durch eine Fusion der Enzyme mit Biotin-Carboxyl-Carrier-Proteinen (BCCP), womit sie an Streptavidin beschichteten paramagnetischen Beads immobilisiert werden konnten. Die An- und



Abb. 6: GS FLX Instrument

Ablieferung der Substrate und Produkte wird durch Diffusion aus einer laminaren Strömung der Reagenzien über der PTP in die Wells erreicht, womit die Reaktionskinetiken der Enzymkaskade an die Diffusionsraten der einzelnen Substanzen angepasst werden können [62].

Um einen Hochdurchsatz zu ermöglichen, müssen Probenvorbereitung und Template-Präparation entsprechend angepasst werden, um Kosten und Zeitaufwand gering zu halten. Eine *in vivo*-Template-Generation analog zum Sanger-Sequencing war nicht möglich. Die von 454 verwendete Template-Generation basiert auf Emulsion-PCR (emPCR) [67,68]. Durch den Einsatz von grenzflächen-aktiven Substanzen [62], die bei der Herstellung von Sprengmitteln eingesetzt werden [69], konnte die Effektivität und Thermostabilität der emPCR maßgeblich verbessert werden. In der emPCR kann Template-DNA *in vitro* im Hochdurchsatz klonal amplifiziert werden [70], weil jeder DNA-Template-Strang in einem einzelnen diskreten „Reaktionstropfen“ amplifiziert wird. Dadurch wird ein Bias durch unterschiedliche Effizienz der Amplifikation [71] und variierendem G+C-Gehalt [72] der Template-DNA weitestgehend vermieden.

1.4.2 Ablauf des 454-Sequenzierens

Der Ablauf des Sequenzierens wird an dem aktuellen Protokoll (Titanium-Version; Lib-L-Kits) demonstriert. Um eine DNA Sequenzierung durchzuführen, muss diese Adaptersequenzen besitzen. Die Adaptersequenzen sind für den Ablauf -die Pyrosequenzierung benötigt einen

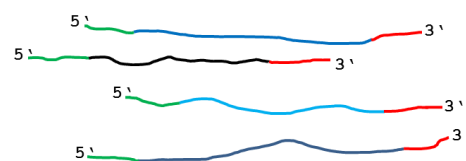


Abb. 7: 454 Library mit Adaptersequenzen

Primer, da durch eine Polymerase dNTPs eingebaut werden, und die unterschiedlichen Aufreinigungsschritte unerlässlich. Eine 454-Library muss am 5'-Anfang Adapter-A-Sequenzen (grün dargestellt in Abb. 7) und am 3'-Ende Adapter-B-Sequenzen besitzen (rot dargestellt in Abb. 7); zur Vereinfachung ist die 454-Library in Abb. 7 einzelsträngig abgebildet, je nach Art der Library-Herstellung kann sie aber auch doppelsträngig sein. Die unterschiedlichen Library Typen werden in Kapitel 1.4.4 vorgestellt.

1.4.2.1 emPCR

Zu der Library werden neben den Komponenten des PCR-Mixes sogenannte Capture-Beads gegeben (Abb. 8), die auf ihrer Oberfläche Oligonukleotide tragen, welche komplementär zu den Adapter-B-Sequenzen der 3'-Enden der Librarys sind. Die Oligonukleotide sind über Hexa-Ethylen-Glycol-Spacer mit den Capture-Beads verbunden; die Basis für die Capture-Beads ist Sepharose [62]. Ist die Library doppelsträngig und sind beide DNA-

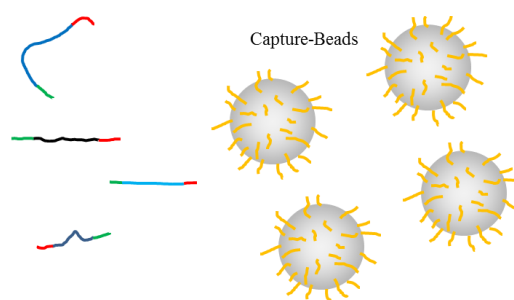


Abb. 8: Capture-Beads

Stränge „aktiv“ bei der Sequenzierung (vergleiche Kapitel 1.4.4), so wird die Library vorher denaturiert. Die im PCR-Mix verwendete DNA-Konzentration der Library ist kritisch; sie muss so gewählt werden, dass idealerweise ein effektives DNA-Library-Molekül auf einen Capture-Bead kommt (Abb. 9), damit während der emPCR auch nur ein DNA-Molekül am Capture-Bead amplifiziert wird (monoklonale Amplifikation). Effektives DNA-Molekül bedeutet, dass es amplifizierbar ist und beide Adaptersequenzen enthält. Kommen zu wenig effektive Library-Moleküle auf die Capture-Beads, so ist die Amplifikation ineffizient; die Ausbeute nach der PCR ist zu klein für einen Sequencing-Run. Kommen dagegen zu viele effektive Library-Moleküle auf die Capture-Beads, so findet keine monoklonale Amplifikation mehr an den Beads statt; der Anteil an Beads, an denen mehr als nur eine Library-Sequenz amplifiziert wird, steigt. Diese Capture-Beads stören das Sequenzieren, da deren Sequenzen nicht bestimmt werden können und verworfen werden.

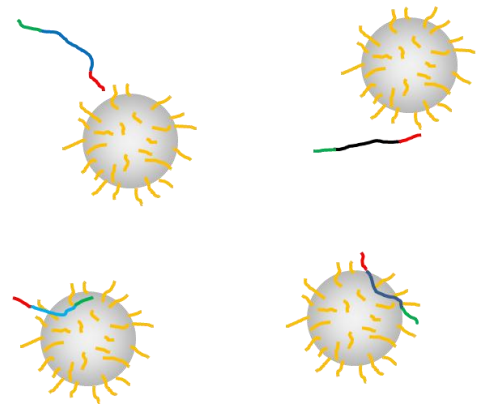


Abb. 9: Ein effektiver DNA-Strang pro Capture-Bead.

Die Einstellung des richtigen Verhältnisses zwischen Library-Molekülen und Capturebeads muss experimentell ermittelt werden (vergleiche Kapitel 1.4.5). Das Verhältnis von Library-Molekülen zu Capture-Beads nennt man cpb-Zahl, „Copy Per Bead“ - Anzahl an DNA-Molekülen pro Capture-Beads.

Nach der Zugabe des Emulsionsöls zum Gemisch aus DNA, PCR-Mix und Capture-Beads (Abb. 10) und Homogenisierung entsteht eine hitzestabile Wasser-in-Öl-Emulsion, die aus PCR-„Mikro-Reaktionsräumen“ mit einem Durchmesser von ca. 50 bis 100 µm besteht. In diesen Mikro-Reaktionsräumen oder Micellen sind alle Komponenten für die PCR enthalten und idealerweise nur ein effektives DNA-Library-Molekül und ein Capture-Bead, so dass eine monoklonale Amplifikation in den Micellen stattfinden kann. In der Emulsion sind diese Mikroreaktionsräume durch das Öl voneinander getrennt, so dass sie sich nicht gegenseitig beeinflussen; sie „sehen“ sich nicht, wodurch es möglich ist, eine Vielzahl von Library-Molekülen mit unterschiedlichen Insertsequenzen in einem PCR-Tube gleichzeitig zu amplifizieren.

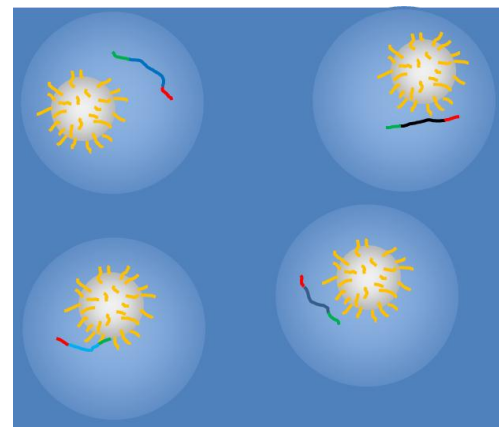


Abb. 10: Ausbildung der Emulsion

Während der emPCR kommt es zum Annealing (Abb. 11) des 3'-Endes der Library (Adapter-B-Sequenz) mit dem komplementären Oligonukleotid des Capture-

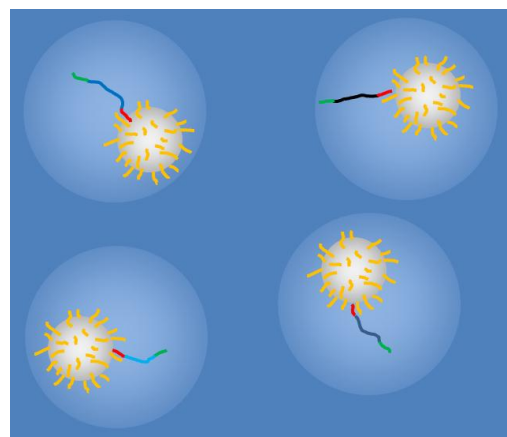


Abb. 11: Library-Capture

Beads, genannt Library-Capture. Das Oligonukleotid dient als Primer für die PCR, es findet eine Elongation statt, so dass der an Hand des Library-Moleküls aufgebaute Gegenstrang dauerhaft mit dem Capture-Bead verbunden ist (Abb. 12).

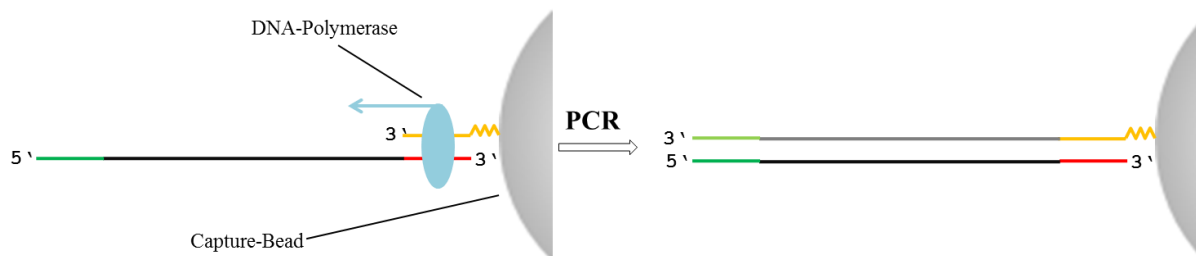


Abb. 12: PCR an Capture-Beads

Da in dem PCR-Mix auch Primer mit den Sequenzen von Adapter-A und der komplementären Sequenz zu Adapter-B vorhanden sind, kann die Vervielfältigung während der PCR auch „außerhalb“ der Capture-Beads stattfinden (Abb. 13). Die so entstandenen Amplifikationsprodukte besitzen die Sequenz des Library-Moleküls, so dass diese in weiteren PCR-Schritten wieder mit den Sequenzen der Capture-Beads Annealen können. Damit ist eine exponentielle Vervielfältigung während der emPCR gewährleistet.

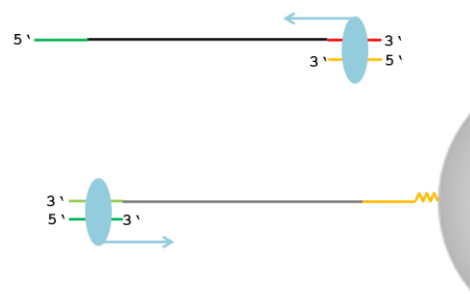


Abb. 13: PCR außerhalb der Capture-Beads

Das Ergebnis der emPCR sind Capture-Beads, an denen idealerweise eine monoklonale Amplifikation stattgefunden hat (Abb. 14). Als Ergebnis der emPCR sind ca. $1-5 \times 10^7$ Kopien der Library pro Bead zu erwarten.

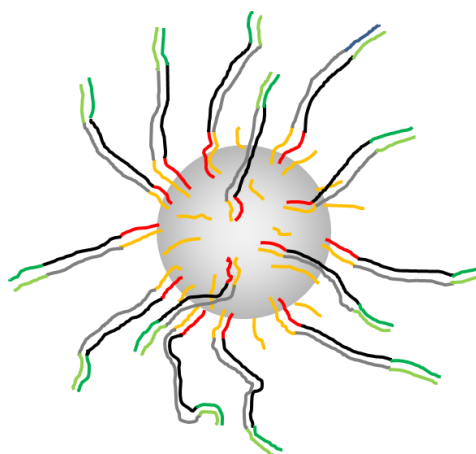


Abb. 14: Capture-Bead mit PCR-Produkten

1.4.2.2 Beadrecovery

Um die Capture-Beads für das Sequenzieren bereit zu machen, müssen diese einige Aufreinigungsschritte durchlaufen, die unter dem Begriff „Beadrecovery“ zusammengefasst werden. Die Amplifikationsprodukte tragenden Capture-Beads müssen von dem Öl der emPCR, sowie den dort verwendeten Enzymen, Nukleotiden und Puffern befreit werden. Dazu wird die Emulsion in mehreren Schritten aufgebrochen und aufgereinigt; die Amplifikationsprodukte werden einzelsträngig gemacht und die nicht an die Beads gebundenen DNA-Stränge entfernt (Abb. 15). Die Anzahl der so erhaltenen Beads wird gemessen; geteilt durch die Zahl der ursprünglich eingesetzten Capture-Beads ergibt das die sogenannte „Beadrecovery Rate“. Mit ihr lassen sich Rückschlüsse auf die Qualität der vorangegangenen Aufreinigungsschritte ziehen; laut Roche sollen 80-85% der Beads zurückgewonnen werden.

Um die Effizienz des Sequencing zu verbessern, müssen die Capture-Beads entfernt werden, die keine oder nur wenige Amplifikationsprodukte tragen. Dazu kommen die sogenannten „Enrichment-Beads“ zum Einsatz (Abb. 16). Das sind magnetische Streptavidin gecoatete Beads, die über Biotin-Linker verbundene Primer mit den Sequenzen von Adapter-A tragen. So können Capture-Beads über die zu Adapter-A komplementären Sequenzen der 3'-Enden ihrer Amplifikationsprodukte an die Enrichment-Beads binden. Durch Waschschrte in einem Magnetic Particle Collector (MPC) werden die Capture-Beads, die keine oder nur wenig Amplifikationsprodukte tragen, entfernt (Abb. 17), da Capture-Beads mit ausreichend vielen Amplifikationsprodukten über die magnetischen Enrichment-Beads im MPC gehalten werden. Nach Waschschrten werden die Enrichment-Beads von den Capture-Beads durch basische Denaturierung getrennt; übrig bleiben Capture-Beads mit den einzelsträngigen Amplifikationsprodukten (Abb. 18). Am Ende des

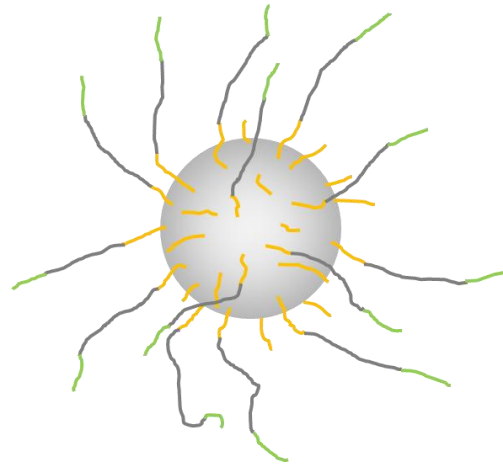


Abb. 15: Capture-Bead mit einzelsträngigem Amplifikationsprodukt

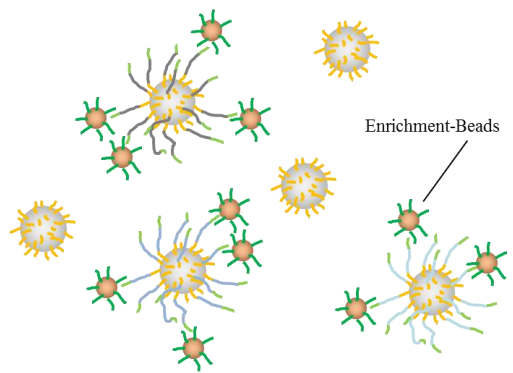


Abb. 16: Binden an Enrichment-Beads

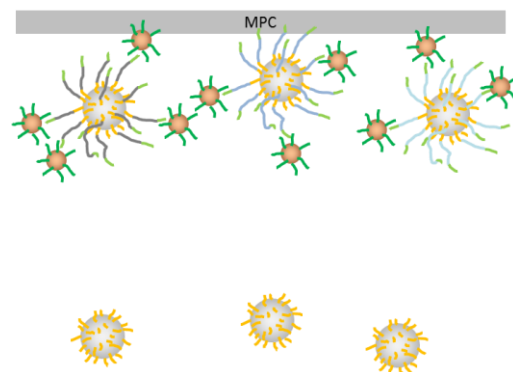


Abb. 17: Waschschrte mit dem MPC

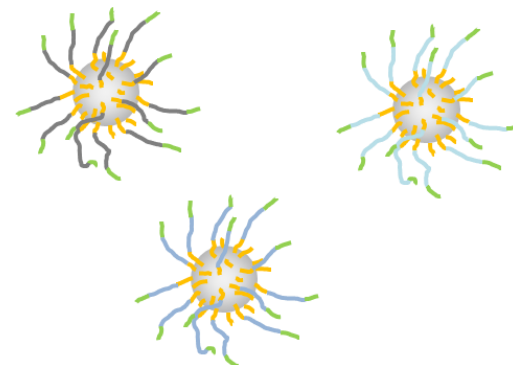


Abb. 18: Capture-Beads nach den Waschschrten im MPC

Beadrecovery wird der Sequencingprimer, der die Adapter-A-Sequenz trägt und mit den entsprechenden komplementären 3'-Enden der einzelsträngigen Amplifikationsprodukte annealen kann, dazugegeben (Abb. 19). Die Anzahl an so erhaltenen Beads wird wieder durch Messung bestimmt; sie wird mit der Anzahl an Beads, die man nach den ersten Aufreinigungsschritten erhalten hatte (Abb. 15), ins

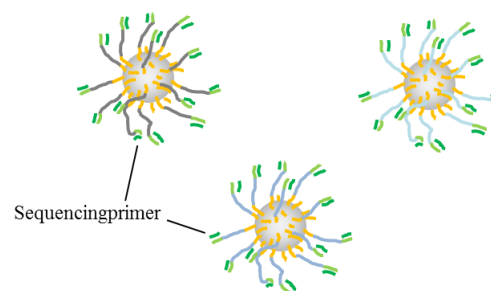


Abb. 19: Ende des Beadrecovery

Verhältnis gesetzt. Dieses prozentuale Verhältnis nennt man Enrichment-Rate. Niedrige Enrichment-Rates bedeuten, dass möglicherweise zu wenig Beads für einen Sequenzier-Laufstart generiert wurden; hohe Enrichment-Rates weisen darauf hin, dass bei einer größeren Anzahl von Beads keine monoklonale Amplifikation mehr stattgefunden hat, sondern es mehr als ein effektives DNA-Library-Molekül pro Capture-Bead gab. Diese Beads verschlechtern das Ergebnis eines Sequenzierlaufs, da deren Reads verworfen werden. Die Obergrenze für die Enrichment-Rate, bei der ein Sequenzieren von Beads vertretbar ist, wird mit 15-20% angegeben [73,74].

1.4.2.3 Sequencing

Jetzt werden die Capture-Beads in die PTP gebracht. Eine PTP wird mindestens in zwei, maximal in 16 unterschiedliche Regionen aufgeteilt; je mehr Regionen man verwenden will, desto größer ist der Verlust an verwendbarem Platz auf der PTP, da die unterschiedlichen Regionen beim Aufbringen der Capture-Beads in die PTP durch Plastik-Spacer voneinander getrennt sind. Da die Menge an verwendeten Sequenzier-Reagenzien aber immer gleich groß und teuer ist, werden zwei Regionen bevorzugt. Die laut Roche zu erwartenden Ausbeuten pro Region sind in der folgenden Tabelle dargestellt (Tab. 1).

Region Size	PTP Device Size	Regions per PTP Device	Bases / Region (Mb)	Bases per Full PTP Device (Mb)	Reads / Region ($\times 10^3$)
Large	70 x 75 mm	2	180-280	360-560	450-650
Medium	70 x 75 mm	4	60-110	240-440	160-250
M/S	70 x 75 mm	8	30-55	240-440	80-120
Small	70 x 75 mm	16	10-20	160-320	25-40

Tab. 1: Ausbeuten pro Region [75]

Zur Vorbereitung des Sequenzierens werden die im Beadrecovery präparierten Capture-Beads mit einer Polymerase inkubiert, damit sich diese an den doppelsträngigen Bereich (vergleiche Abb. 19) der Capture-Beads anlagern kann. Bei dem verwendeten Enzym handelt es sich um *Bst*-DNA-Polymerase aus *Bacillus stearothermophilus* [76]. *Bst*-DNA-Polymerase verfügt über eine hohe Strang-Verdrängungs-Aktivität [77], ist gut geeignet für die Replikation von DNA mit hohem GC-Gehalt [78], kann geringe Mengen DNA schnell sequenzieren [79] und besitzt eine hohe Prozessivität [80], weshalb sie für das 454-Sequencing ausgewählt wurde [62]. Nacheinander werden durch Zentrifugations-Schritte als erstes Enzyme-Beads, dann zweitens Capture-Beads zusammen mit sogenannten Packing-Beads, dann wieder Enzyme-Beads und als letztes PPI-ase-Beads in die PTP

zentrifugiert, so dass es in einem PTP-Well unterschiedliche Bead-Schichten bzw. -Layer gibt (Abb. 20). Die Enzyme-Beads tragen über BCCP verbunden die Enzyme Luciferase und ATP-Sulfurylase. Die Packing-Beads dienen zum „Verkanten“ der Capture-Beads in dem PTP-Well, damit diese vor dem Herauswaschen bewahrt werden. Die PPI-ase-Beads stellen die oberste Schicht dar; einerseits behindern sie so nicht die Enzymkaskade, auf der anderen Seite können sie überschüssiges herausdiffundierendes

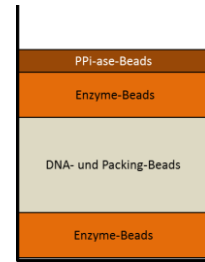


Abb. 20: Bead-Layer

Pyrophosphat abbauen; damit wird eine Kontamination anderer Wells und eine so mögliche falsche Signalerzeugung verhindert. Zusätzlich werden auch noch sogenannte „Control-Beads“ mit auf die PTP gebracht; das sind von Roche hergestellte Beads, deren Sequenz bekannt ist und damit als Qualitätsmerkmal für das Sequencing herangezogen wird. Unten sind vergrößerte Abbildungen von Enzyme-, Capture- und Packing-Beads im PTP-Well (Abb. 21) sowie einer PTP mit allen Bead-Layern dargestellt (Abb. 22); ferner PTPs nach einem erfolgten Sequenzier-Lauf mit 2 (Abb. 23) und 4 Regionen (Abb. 24). Während des Sequenziervorganges werden die Substrate (Luciferin, APS und SSB) und unterschiedliche dNTPs jetzt nacheinander in Zyklen (mit dazwischen liegenden Apyrase-Waschschritten) über die PTP gespült (Flows). Die Polymerasen beginnen an den doppelsträngigen Bereichen der Amplifikationsprodukte den Strangaufbau anhand der Matrize (Abb. 25 auf Seite 15), sofern entsprechend komplementäre dNTPs vorhanden sind. Die Enzymkaskade kann bei erfolgreichem Einbau Lichtblitze erzeugen (Abb. 5 auf Seite 7). Der Gegenstrang wird solange aufgebaut, bis entweder die Polymerase am Ende angekommen ist oder vorher abbricht, weshalb nicht immer bis Adapter-B durchgelesen werden kann (Abb. 26 auf Seite 15).

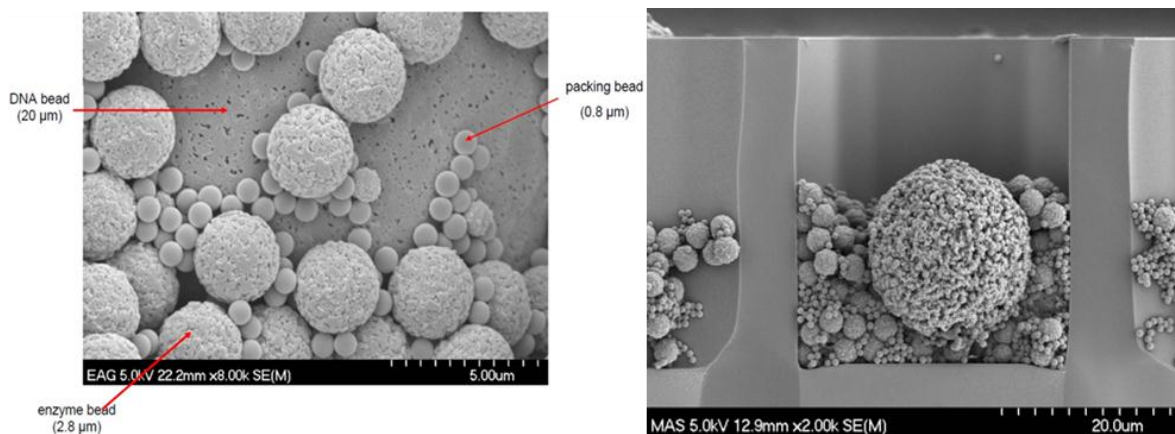


Abb. 21: Enzyme-, Capture- und Packing-Beads in der PTP (www.roche-applied-science.com)

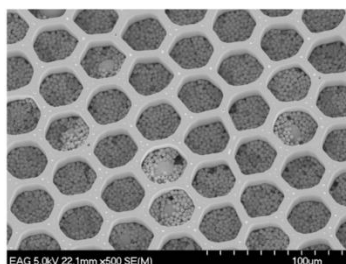


Abb. 22: Vier Bead-Layer (www.roche-applied-science.com)



Abb. 23: 2 Regionen



Abb. 24: 4 Regionen

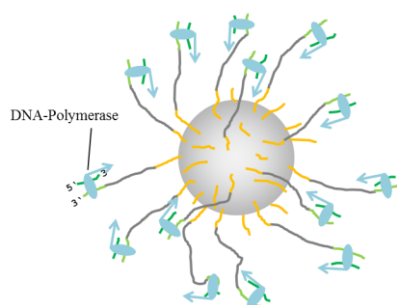


Abb. 25: Sequencing an den Beads

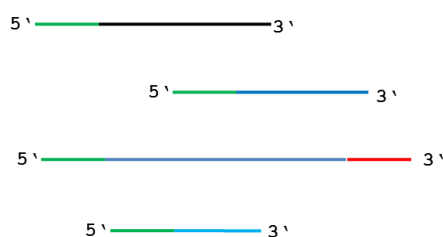


Abb. 26: Ergebnis des Sequencings

1.4.3 Data-Processing

Während des Sequenzierlaufs wird von der PTP bei jedem dNTP-Flow ein Bild von der CCD-Kamera aufgenommen; diese Bilder stellen die Rohdaten (Rawdata) dar und werden in zwei Schritten ausgewertet bzw. prozessiert.

1.4.3.1 Imageprocessing

In einem ersten „Imageprocessing“ genannten Schritt werden die Wells auf der PTP bestimmt, die überhaupt einen Lichtblitz von sich geben. Über den ersten signalerzeugenden Flow werden die Regionen der PTP definiert. Zudem werden die Raw-Wells definiert, also Wells in der PTP, die einen DNA-tragenden Capture-Bead enthalten. Das geschieht durch Aufsummieren aller Orte auf der PTP, die während der ersten Flows ein Lichtsignal von sich geben (sogenannte „Key-Flows“, die Flows, die zur Bestimmung des Key notwendig sind –vergleiche Kapitel 1.4.4 auf Seite 17). Außerdem wird die Hintergrundstrahlung für jedes dNTP während der Key-Flows ermittelt und von den Signalen subtrahiert

1.4.3.2 Signalprocessing

Die im Imageprocessing erhaltenen Daten werden in einem zweiten Schritt, dem „Signalprocessing“, letztendlich in Sequenzen umgewandelt. Es finden Normalisierungen, Korrekturen, und Filterschritte statt. Reads, die diese Schritte passieren, werden in sogenannte „Flowgrams“ für jedes Well umgeschrieben. Für jeden Read generiert das Signalprocessing Basecalls mit dazugehörigen Quality Scores, und speichert diese Daten als „Standard Flowgram Format“ (SFF). Es finden folgende Schritte statt:

Da der Einbau der unterschiedlichen dNTPs zu unterschiedlich starken Signalen führt, müssen diese untereinander normalisiert werden. Gibt ein Well besonders starke Signale von sich, weil an dem Capture-Bead sehr viele Amplifikationsprodukte sitzen oder weil es homopolymere Bereiche gibt (z.B. TCAAAAAACGC), so kann es zur Überstrahlung benachbarter Wells kommen, was korrigiert wird. Beim Fluss der Substrate über die PTP kann es zu lokalen Unterschieden kommen, besonders am Ende der Platte, so dass dort die Signale schwächer sind, was ebenfalls behoben wird.

Wie beim klassischen Pyrosequencing, kommt es auch beim 454-Sequencing zu kleinen Synchronitätsverlusten an den Capture-Beads [81], sogenannte Out-Of-Phase-Fehler müssen behoben werden. Verbleiben Spuren von dNTPs nach dem Apyrasewaschschritt im Well, so kann es zu vorzeitigem Einbau des Nukleotides an einigen Strängen der Amplifikationsprodukte am Capture-Bead kommen; dieser „Carry Forward“ genannte Effekt tritt bei ca. 2% der Stränge per Bead auf und muss erkannt und korrigiert werden. Ein zweiter Out-Of-Phase-Fehler, genannt „Incomplete Extension“ muss ebenfalls erkannt und behoben werden: an einigen Strängen am Capture-Bead werden manchmal keine entsprechenden dNTPs eingebaut. Gründe hierfür können unterschiedliche Reaktivitäten sein (bei dTTP-Flows tritt das Phänomen öfter auf), oder geringere Substratkonzentrationen, meist am Ende der PTP.

Während des über 8 Stunden andauernden Sequencing-Runs kommt es am Ende zu einem generellen Abfall der Signalstärke über die gesamte PTP, was ebenfalls korrigiert werden muss. Außerdem werden sogenannte Ghost-Wells herausgerechnet, indem die Signalstärke pro Base kalkuliert und mit der anderer Wells verglichen wird.

Die so korrigierten Wells, die nach diesen Schritten übrig bleiben, gelten als echte Sequenzen, allerdings noch von unterschiedlicher Qualität. Um letztendlich Sequenzen hoher Qualität zu erhalten, folgen weitere Signal-Qualitäts-Filter und ein Kürzen mancher Reads (Read-Quality- und Read-Trimming-Filter). Folgende Quality-Filter kommen zum Einsatz:

Der „Key-Pass“-Filter entfernt die Reads, die weder als Capture-Bead-Amplifikationsprodukt zu identifizieren sind, noch Kontroll-DNA der Control-Beads sind. Dies geschieht über die Bestimmung einer 4 Basen langen Sequenz, genannt „Key“, die direkt auf die Adapter-A-Sequenz folgt (näheres folgt in Abschnitt 1.4.4). Sequenzen, die diesen Filter erfolgreich passieren, werden als „Key-Passed-Wells“ herausgegeben.

Der „Dots“-Filter entfernt Reads mit zu vielen negativen Flows, d.h. Flows, bei denen kein Basensignal erkannt werden (kein Lichtblitz). Als „Dot“ werden drei negative Flows hintereinander angesehen. Es werden die Reads entfernt, die entweder zu kurz sind (Dot nach weniger als 84 Flows, das entspricht 30-50bp) oder zwar länger als 84 Flows sind, bei denen aber mehr als 5% der Flows Dots sind. Die so entfernten Reads werden als „Failed Dot“ bezeichnet.

Der „Mixed“-Filter dagegen entfernt Reads, bei denen ein vermeintlich zu großer Nukleotid-Einbau während der Flows stattgefunden hat. Außerdem werden die Reads entfernt, deren normalisiertes Signal bei einem Flow weder stark genug für ein „echtes“ Signal noch schwach genug ist, um als negativer Flow zu gelten. Der „Mixed“-Filter entfernt also Reads, die aus Wells stammen, die höchstwahrscheinlich entweder mehr als ein Capture-Bead enthalten, oder ein Capture-Bead enthalten, das aber mehr als ein Amplifikationsprodukt trägt (keine monoklonale Amplifikation), oder von dem Signal eines benachbarten Wells zu stark überstrahlt wird. Die so entfernten Reads werden als „Failed Mixed“ bezeichnet.

Alle Reads, die diese Quality-Filter passieren, werden noch durch folgende Trimming-Filter gekürzt bzw. verworfen:

Der „Signal Intensity-Filter“ verkürzt das 3'-Ende der Reads derart (Trimming), dass weniger als 3% der Flows grenzwertige Signalintensitäten aufweisen (Signalintensitäten zwischen 0,5 und 0,7 auf einer Skala von 0 bis 1). Grenzwertige Signalintensitäten, also schlechtere und schwächere Signale kommen am Ende eines Runs durch den Gesamtabfall der Signalintensität während des Runs und Akkumulation der Out-Of-Phase-Fehler zustande. Reads, die nach dem Trimming kürzer als 40 bp sind (Standard-Einstellungen), werden als „Short-Quality“ herausgefiltert.

Der Primer-Filter entfernt kurze Reads, bei denen die Adapter-B-Sequenz identifiziert wird.

Der TrimBack-Valley-Filter beschneidet oder entfernt Reads mit zu starker intermediärer Signalintensität. Zwischen dem Einbau von einer und zwei gleichen Basen (also beispielsweise zwischen einem A und TT) oder zwischen zwei gleichen und drei gleichen Basen (z.B. zwischen CC und GGG) sollte die Signalstärke idealerweise auf Null abfallen; ist dies nicht der Fall, werden diese Signale als intermediäre Signale identifiziert (Borderline-Valley-Flow). Ein erhöhter Anteil an intermediären Signalen bedeutet eine schlechtere Read-Qualität, die Fehlerquote bei der Bestimmung der genauen Basenzahl von homopolymeren Bereichen (also AA, TTT, usw.) steigt. Ein Read wird entfernt, wenn mehr als vier Borderline-Valley-Flows innerhalb der ersten 320 Flows auftreten; die entfernten Reads werden ebenfalls als „Short-quality“ bezeichnet [82].

1.4.4 Library-Arten

Jede 454-Library besteht immer aus einer Adapter-A-Sequenz am 5'-Anfang, gefolgt von einer Key-Sequenz, dann dem Insert, also der DNA-Sequenz, die bestimmt werden soll, und am 3'-Ende einer Adapter-B-Sequenz (vor der B-Sequenz ist meistens auch noch eine revers-komplementäre Key-Sequenz vorhanden, da manche Library-Arten auch revers sequenziert werden können). Libraries, die mit dem aktuellen Titanium oder Titanium-plus-System verwendet werden sollen, tragen andere Adapter-Sequenzen als die früheren Libraries zu „Standard-FLX“-Zeiten (vergleiche Kapitel 1.4.1). Das war bei der Umstellung von Standard-FLX auf Titanium insofern ärgerlich, als dass die alten Libraries nun nicht mehr auf dem neuen Titanium-System verwendet werden konnten. Die Sequenzen einer Titanium-Library sehen prinzipiell wie folgt aus:

5' **CCATCTCATCCCTGCGTGTCTCCGACTCAG**-Insert-CTGAG**ACTGCCAAGGCACACAGGGGATAGG**3'

Adapter-A ist in grün dargestellt, Adapter-B rot, der Key ist unterstrichen bzw. der revers komplementäre Teil kursiv angegeben. Bis auf eine Ausnahme, den Rapid-Libraries, sehen alle Adaptersequenzen so aus. Die Funktion von Adapter-A und -B-Sequenzen ist in den vorangegangenen Kapiteln erklärt worden. Die vier Basen lange Key-Sequenz wird benötigt, um Insertsequenzen von Kontroll-DNA-Sequenzen zu unterscheiden, da Kontroll-DNA-Sequenzen andere Keys tragen.

Um mehr als eine Library pro Region sequenzieren zu können, weil beispielsweise die Anzahl an Reads, die bei einem 454-Run auf einer Region erhalten werden, viel größer ist als die Zahl der benötigten Sequenzen, können sogenannte Multiplex Identifiers (MIDs) eingeführt werden. Sie stellen einen DNA-Barcode dar, anhand dessen die erzeugten Sequenzen den unterschiedlichen Librarys zugeordnet werden können. Roche verwendet dafür 10 Basen lange Sequenzen als MIDs, die auf die Key-Sequenzen folgen:

5' CCATCTCATCCCTGCGTGTCTCCGACTCAGTGATACGTCT-Insert-CTGAGACTGCCAAGGCACACAGGGGATAGG3'

Das MID (in diesem Beispiel MID Nr.11) ist gelb unterlegt. Die Sequenzen der insgesamt 150 von Roche bereitgestellten MIDs sind so gewählt, dass sie möglichst wenig Flows benötigen, akkurat in der Oligosynthese hergestellt werden können, möglichst keine monomeren Repeats enthalten (z.B. TTT), selbst bei zwei Fehlern (Missmatches) noch eindeutig identifizierbar sind, und Nonsense ergeben, wenn zwei verschiedene MID-Sequenzen an dem gleichen Capture-Bead amplifiziert wurden [83,84].

Entscheidend für die Qualität der Librarys und erfolgreicher Sequenzierungen ist die geforderte Größenverteilung. Generell lassen sich Library-Produkte, die größer als 1kb sind, schlecht in der emPCR amplifizieren [74], weshalb bei Fragmentierungsschritten darauf geachtet werden muss, dass die Fragmentierungsprodukte nicht zu groß ausfallen. Ebenfalls müssen kleinere oder unerwünschte Produkte beseitigt werden, da sie sonst zu kurzen Sequenzierprodukten führen. Dies wird mit Gelausschnitten oder der sogenannten „Solid-phase reversible immobilization“-Methode (SPRI) erreicht.

Bei den von Roche vorgegebenen Möglichkeiten der Library-Herstellung gibt es prinzipiell zwei Wege, um die oben beschriebenen Adaptersequenzen zu der Insert-DNA hinzuzufügen; entweder über eine Random-Ligation oder über das Heranamplifizieren der Adaptersequenzen mittels PCR. In den folgenden Abschnitten wird dargestellt, welche unterschiedlichen Arten von Librarys es beim 454-Pyrosequencing gibt.

1.4.4.1 Titanium General Library

Dies war die erste Anwendung auf dem Titanium-System. Sie zeichnet sich durch Zufalls-Ligation des Adapters an die Inserts aus (Random-Ligation) mit einer daraus resultierenden ungerichteten Sequenzierung. Die benötigten DNA-Ausgangsmengen liegen mit und ohne Fragmentierung bei 5 µg und 1 µg. Die Library ist am Ende der Herstellung einzelsträngig und trägt den TCAG-Key. Das Prinzip der Herstellung ist in Abb. 27 auf der folgenden Seite dargestellt.

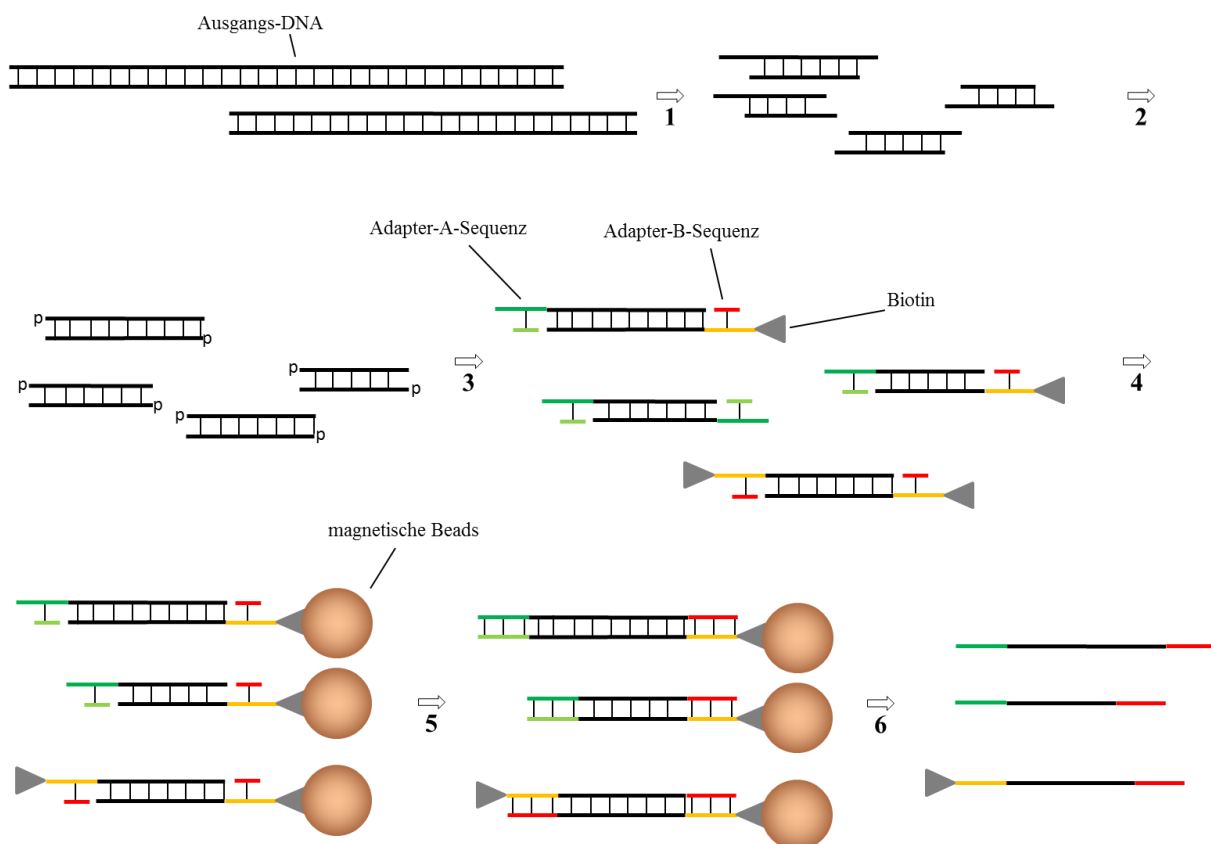


Abb. 27: Herstellung einer Titanium General Library

Die Ausgangs-DNA (genomische DNA, BACs, Fosmide, usw.) wird durch Nebulisierung in kleinere Produkte fragmentiert (1). Nach einer Größenselektion von Fragmenten zwischen 500-800 bp findet ein Endpolishing statt. Die teilweise ausgefransteten Enden der Fragmentationsprodukte werden mittels Polymerasen zu Blunt-Enden repariert und eine Phosphonukleotid-Kinase hängt Phosphatgruppen an die 5'-Enden der Fragmentationsprodukte heran (2). Will man eine Library erstellen, die nicht fragmentiert werden muss, weil die Inserts der DNA schon entsprechend klein sind (z.B. PCR-Produkte oder kurze cDNAs), dann wird die Fragmentierung übersprungen und man steigt mit dem Endpolishing ein. Danach erfolgt die Ligation der Adapter, die die Adapter-A- und -B-Sequenzen sowie den Key tragen; Adapter-B trägt auch ein Biotin-Tag. Will man der Library ein MID geben, so werden Adapter verwendet, die nach Adapter-A-Sequenz und Key das entsprechende 10 Basen lange MID enthalten. Die Adapter sind so aufgebaut, dass sie nur mit ihren 3'-Key-Enden über eine Blunt-End-Ligation mit den 5'-phosphorierten-Enden der Inserts ligiert werden können, da die Adapter auf der anderen Seite 5'-overhangs besitzen. Die Adapter tragen keine Phosphatgruppen an ihren 5'-Enden, um eine Adapter-dimer-Bildung über die Key- oder MID- (bei Adapter mit MIDs) tragenden Blunt-End-Bereiche zu minimieren. Daher findet die Blunt-End-Ligation auch nur an einem Strang statt, am anderen verbleibt ein Nick (3). Die Adapter werden in einem Verhältnis von 15:1 zu den Fragmenten gegeben, um die Bildung von Fragment-Konkatemeren bei der Ligation zu verhindern. Nach einer SPRI-Aufreinigung bindet nun die Library über das Biotin-Tag von Adapter-B an Streptavidin beschichtete magnetische Beads (4). Diese werden im MPC immobilisiert und gewaschen, um mögliche unligierte Produkte, Konkatemere und restliche Adapterdimere zu entfernen. Anschließend findet eine Reparatur der Nicks zwischen Adapter und Insert durch eine Polymerase statt; die Adapter sind nun über beide Stränge mit den Inserts verbunden (5). Am Ende wird die Library von den magnetischen Beads über basische Denaturierung isoliert (6) und man erhält ein einzelsträngiges Produkt.

Eine fragmentierte General Library soll zwischen 500 und 800 bp liegen, wobei jeweils weniger als 10% der Library kleiner als 350 bp und größer als 1000 bp sein sollen. In Abb. 28 ist das Konzentrationsprofil einer General Library auf einem 7500 DNA Chip dargestellt [85].

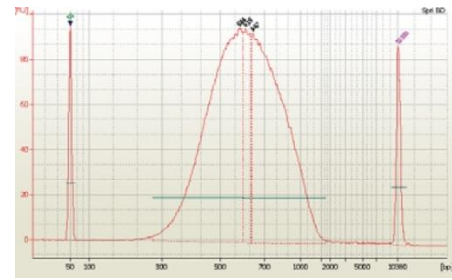


Abb. 28: Profil einer General Library [85]

1.4.4.2 Titanium Paired End Library

Da die mit dem 454-Pyrosequencing erzeugten Reads kürzer als Sangersequenzen sind, kann es zu Problemen beim Assemblieren der Sequenzdaten kommen, z.B. bei *de novo* Sequenzierung von Genomen. Um das Scaffolding (das Zusammenbauen) von Contigs (contiguous sequences) bei schwierigen Sequenzbereichen mit geringem Coverage zu ermöglichen, werden Paired End Libraries erstellt, durch welche Sequenzinformationen zu langen DNA-Bereichen gewonnen werden können [86]. In Abb. 29 ist das Prinzip der Herstellung einer Paired End Library dargestellt [87-89].

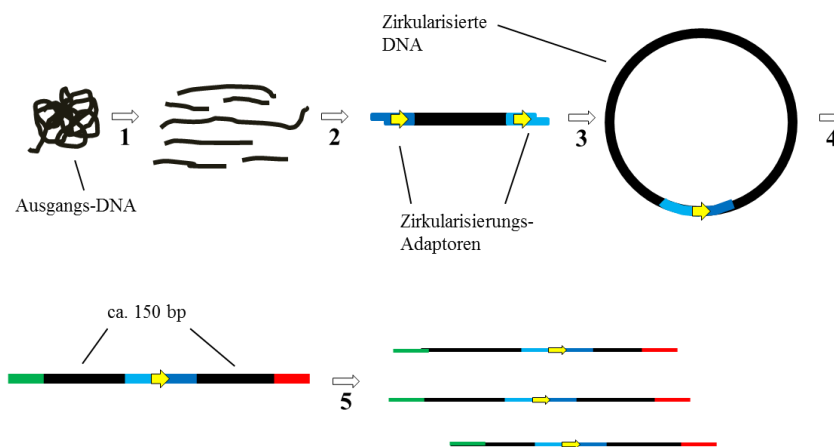


Abb. 29: Herstellung einer Titanium Paired End Library

Die Ausgangs-DNA (meistens genomische DNA) wird in den bevorzugten Größenbereich fragmentiert (1). Die Protokolle von Roche beinhalten Größen von 20 kb, 8 kb und 3 kb (Abstand zwischen den zu bestimmenden Enden), wobei man für eine 20 kb-Library 30 µg, für 8 kb 15 µg und für 3 kb 5 µg an Ausgangs-DNA-Menge benötigt. Die DNA unterläuft ein Endpolishing analog zum General Library Protokoll, bevor eine Ligation der Zirkularisierungs-Adapter (blau dargestellt) stattfindet (2). Diese tragen loxP-Sites (gelb dargestellt), so dass das *Cre-lox*-Rekombinase-System angewandt werden kann. Die Zirkularisierungs-Adapter sind analog zu den A- und B-Adaptoren aufgebaut. Sie haben auf der einen Seite Overhangs und auf der anderen kann nur ein Strang mit der doppelsträngigen Insert-DNA ligiert werden (vergleiche Abb. 27). Es folgt eine Größenselektion und anschließend eine Nick-Reparatur. Nun findet mittels *Cre*-Rekombinase eine Zirkularisierung statt (3). Mit diesen Zirkularisierungs-Produkten wird jetzt analog wie bei der Herstellung einer General Library verfahren (4); es folgen Fragmentierung, Endpolishing, und Ligation von Adapter-A (grün dargestellt) und -B (rot dargestellt). Die Ligationsprodukte unterlaufen jetzt noch einen Amplifikationsschritt per PCR, um die erhaltene DNA-Menge zu vergrößern. Nach einer Größenselektion der Amplifikationsprodukte mit SPRI-Beads, werden einzelsträngige Paired End Library Produkte isoliert (5), ebenfalls analog der General Library-Herstellung. Die Sequenzen der Fragmentenden belaufen sich auf durchschnittlich 150 bp, sie sind voneinander durch die Zirkularisierungs-Adaptersequenzen getrennt.

1.4.4.3 Titanium Rapid Library

Ende 2009 brachte Roche die Rapid Library auf den Markt, sie ersetzt die Titanium General Library. Rapid Libraries zeichnen sich wie General Libraries durch Randomligation und daraus resultierender ungerichteter Sequenzierung aus; allerdings benötigt man zur Herstellung einer Rapid Library nur 500 ng DNA (bei Fragmentierung) und das Protokoll ist aufgrund weniger Arbeitsschritte wesentlich schneller. Am Ende der Library-Herstellung erhält man ein doppelsträngiges Produkt, das aber, da beide Stränge sequenziert werden, vor der emPCR denaturiert werden muss (vergleiche Kapitel 1.4.2.1). Rapid Libraries tragen als Key die Sequenz GACT. In Abb. 30 ist das Prinzip der Herstellung einer Rapid Library erläutert.

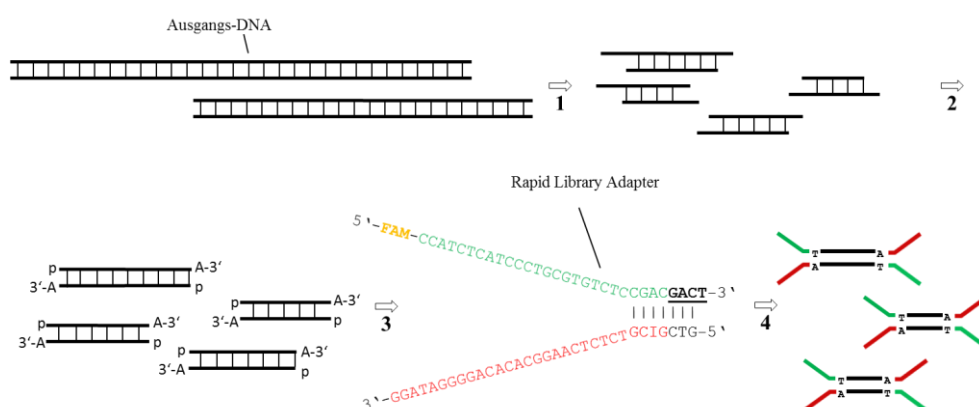


Abb. 30: Herstellung einer Rapid Library

Die Ausgangs-DNA (genomische DNA, BACs, Fosmide) wird durch Nebulisierung in kleinere Produkte fragmentiert (1). Analog der General Library findet auch hier wieder ein Endpolishing statt und eine Phosphonukleotid-Kinase hängt Phosphatgruppen an die 5'-Enden der Fragmentationsprodukte heran. Im Gegensatz zur General Library verwendet man beim Endpolishing u.a. eine Taq-Polymerase, so dass A-Overhangs an den 3'-Enden entstehen (2). Will man eine Library erstellen, die nicht fragmentiert werden muss, dann steigt man mit dem Endpolishing ein. Die A-Overhangs ermöglichen es, dass eine Ligation mit den Y-förmigen Rapid Library Adapter stattfinden kann (3). Die Adapter sind so aufgebaut, dass sie sowohl die Adapter-A-Sequenz (grün) tragen, als auch eine leicht veränderte Adapter-B-Sequenz (rot); diese ist aber immer noch kompatibel mit den emPCR-, Beadrecovery- und Sequencing-Kit, die für General Library oder Paired End Libraries verwendet werden. Der Key ist unterstrichen, wobei das T des Key als Gegenbase für die A-Overhangs der Produkte aus dem Endpolishing fungiert. Will man der Library ein MID geben, so werden Adapter verwendet, die nach dem Key ein 10 bp langes MID besitzen, wobei nach der MID-Sequenz am 3'-Ende das benötigte T als Overhang zu finden ist. Am 5'-Ende der Adapter-A-Sequenz befindet sich ein Fluoreszenzfarbstoff (FAM), über den eine Quantifizierung der Library ermöglicht wird. Nach der Ligation findet nur noch eine Größenselektion über SPRI statt, und man erhält doppelsträngige Libraryprodukte, wobei beide Stränge sequenziert werden, da jeder Strang am 5'-Anfang die Adapter-A-Sequenz und am 3'-Ende die Adapter-B-Sequenz trägt (4).

Profil und Kriterien der Größenverteilung einer Rapid Library entsprechen denen der General Library (vergleiche Abb. 28), nur dass bei der Rapid Library der Größenbereich zwischen 600-900 bp liegt, also 100 Basen länger ist [90].

Auch das Sequenzieren von cDNAs kann mit dem Rapid Library Protokoll stattfinden, dafür wird die RNA fragmentiert und die cDNA-Synthese wird über Random-Primer durchgeführt. Die so erhaltenen Produkte werden dann für die Generation einer Rapid Library benutzt [91]. Es ist dementsprechend also auch eine ungerichtete Shotgun-Sequenzierung.

Seit Anfang 2012 besteht die Möglichkeit, Paired End Libraries mit Rapid Library-Adaptoren zu versehen [92,93]. Mit der zirkularisierten DNA (vergleiche Abb. 29 auf Seite 20) wird wie bei der Herstellung einer normalen Rapid Library verfahren, wobei die im ursprünglichen Paired End-Protokoll enthaltenen Amplifikations- und Aufreinigungsschritte am Ende der Library-Herstellung beibehalten wurden; auch die relativen großen Eingangsmengen an DNA sind gleich geblieben.

1.4.4.4 Titanium Amplicon Library

Ein Amplicon ist ein spezifisches Amplifikations-Produkt. 454-Amplicon Libraries sind dadurch gekennzeichnet, dass die benötigten 454-Adapter-A- und -B-Sequenzen über eine PCR an die Insert-DNA eingeführt werden. Dabei muss ein Teil der Sequenz der zu untersuchenden DNA bekannt sein, um als Ansatzpunkt für die PCR dienen zu können. In Abb. 31 ist das Prinzip der Herstellung einer Amplicon Library dargestellt.

Amplicon Libraries stellen als einzige Library-Variante eine Möglichkeit für gerichtetes Sequenzieren dar, da es hierbei keine Random-Adapter-Ligation gibt und durch die Primerwahl die Sequenzrichtung festgelegt werden kann.

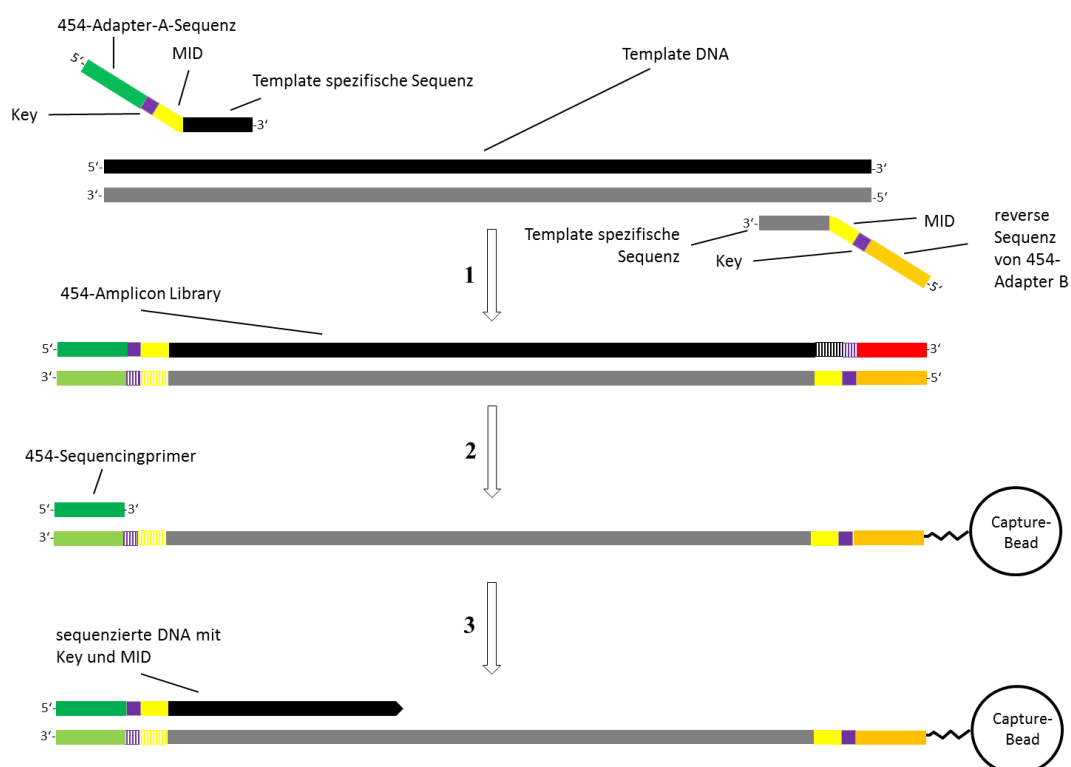


Abb. 31: Prinzip der Herstellung einer Amplicon Library

Die Primer, die bei der Herstellung einer Amplicon Library verwendet werden, tragen an ihren 3'-Enden Template spezifische Sequenzen. Beim Forward-Primer folgen darauf in 5'-Richtung eine MID-Sequenz (gelb), der Key (violett) und die Adapter-A-Sequenz (grün). Beim Revers-Primer folgen auf den Template-spezifischen Bereich in 5'-Richtung eine MID-Sequenz, wenn die Library auch revers sequenziert werden soll, dann der Key, und dann die revers-komplementäre Sequenz der Adapter-B-Sequenz. Mit diesen Primern werden nun durch PCR auf der Template-DNA die 454-Adapter an die Insert- bzw. Template-DNA heranamplifiziert (1), es entsteht eine 454-Amplicon Library. Diese wird dann wie beschrieben in der emPCR an Capture-Beads amplifiziert, aufgereinigt (2) und dann in der PTP sequenziert (3).

Als optimaler Größenbereich wird von Roche eine Amplicon-Länge von 200-600 bp angegeben, was die Adapter-A- und -B-Sequenzen, den Key und die MID's mit einschließt [94]. Die Ausbeuten an Reads pro Region nach einem Sequencing-Run von Amplicon Librarys (Tab. 2) wird von Roche geringer angegeben als die von General- oder Rapid Librarys (vergleiche Tab. 1 auf Seite 13).

Region Size	PTP Device Size	Regions per PTP Device	Reads / Region ($\times 10^3$)
Large	70 x 75 mm	2	~375
Medium	70 x 75 mm	4	~150
M/S	70 x 75 mm	8	~75
Small	70 x 75 mm	16	~20

Tab. 2: Ausbeuten von Amplicon Librarys [94]

Als Adapter-A- und -B-Sequenzen können entweder die gleichen gewählt werden wie für General-, Paired End- und Rapid Library, was Sequenzlängen (einschließlich Key-Sequenz) von 30 Basen bedeutet (vergleiche Abschnitt 1.4.4). Wird ein MID verwendet, so kommen nochmal 10 Basen hinzu. Folglich können die 5'-Overhangs der Primer für die Herstellung einer Amplicon Library also bis zu 40 Basen lang sein. Addiert man dazu die Basenzahl der Template-spezifischen Bereiche am 3'-Ende (ca. 20 bp), so ist der gesamte Primer in etwa 50-60 Basen lang.

Roche bietet auch noch andere Adapter-A- und -B-Sequenzen (welche mit Key 5 Basen kürzer sind, also eine Länge von 25 Basen haben) mit dazugehörigen Kits (Lib-A-Kits) an. Diese Lib-A-Kits sind so gestaltet, dass an der einen Hälfte der Capture-Beads eine Forwardsequenzierung, an der anderen eine Reverssequenzierung stattfindet; es wird also bidirektional sequenziert.

Eine bidirektionale Sequenzierung wird von Roche immer dann empfohlen, wenn durch Forward- und Reverse-Sequenzen eine höhere Konsensusbildung und Genauigkeit erreicht werden kann; dies gilt insbesondere für kürzere Amplicon Librarys, weil die Genauigkeit der Basenbestimmung am Anfang der Reads am höchsten ist bei der 454-Sequenzierung und eine Überlappung von Forward- und Reversereads mindestens 100 bp lang sein soll.

Bei längeren oder stark variablen Amplicon Librarys soll eine unidirektionale Sequenzierung bevorzugt werden, da so ein Deepsequencing aufgrund der maximalen Ausnutzung der PTP ermöglicht wird.

Will man Amplicon Librarys sequenzieren, so kann man ein anderes Signal Processing verwenden, das speziell auf Amplicon Librarys zugeschnitten ist: das Signal Processing for Amplicons. Dieses ist an die geringere Heterogenität der Amplicon Librarys, an die oft kürzeren Sequenzen und deren höheren Signalintensitäten angepasst. Außerdem findet ein erweitertes Screening nach Ghost-Wells statt [82]. Das Signal Processing for Amplicons wird vorallendingen für kurze Amplicon Librarys oder bidirektionales Sequenzieren empfohlen; für unidirektionales Sequenzieren oder längere Amplicon Librarys wird eher zum normalen Signal Processing geraten [94].

1.4.5 Berechnung, cpb-Zahl und Titration

Um einen erfolgreichen Sequenzierlauf durchführen zu können, muss sichergestellt werden, dass in der vorangehenden emPCR genau ein 454-Library-DNA-Molekül an einem Capture-Bead amplifiziert wird. Ist zu wenig DNA vorhanden, so ist die Anzahl an Capture-Beads ohne Amplifikationsprodukt zu groß, und ein Sequenzierlaufstart ist aufgrund von zu wenig DNA tragenden Capture-Beads nicht möglich; kommen zu viele Library-DNA-Moleküle auf einen Capture-Bead, dann verschlechtert dies das Ergebnis eines Sequenzierlaufs, weil Capture-Beads, die mehr als ein Amplifikationsprodukt tragen, beim Filtering aussortiert werden (vergleiche Kapitel 1.4.3.2). Man muss also ein ideales Verhältnis von Library zu Capture-Bead-Anzahl finden. Das Verhältnis von Library zu Capture-Beads wird cpb-Zahl genannt.

Mit Ausnahme der Rapid Librarys, bei denen man aufgrund des FAM-Moleküls an der Adapter-A-Sequenz direkt die Konzentration in Molekülen pro Volumen bestimmen kann, muss bei allen anderen Librarys für das Herausfinden des idealen Verhältnisses erst die Molekülkonzentration errechnet werden. Dafür müssen Konzentration (Masse pro Volumen) und Größenverteilung bekannt sein; die Umrechnung daraus in Konzentration in Molekülen pro μl ergibt sich aus folgender Formel:

$$\text{Konzentration[Moleküle}/\mu\text{l}] = \frac{\text{Library Konzentration[ng}/\mu\text{l}] \times 6,022 \times 10^{23}}{328,3 \times 10^9 \times \text{durchschnittliche Fragmentlänge[Nukleotiden]}}$$

Wobei $6,022 \times 10^{23}$ die Avogadro-Konstante [Moleküle/mol] und 328,3 das durchschnittliche Gewicht eines Nukleotids [g/mol] ist. Wenn die Library doppelsträngig ist, aber nur ein Strang „454-aktiv“ sequenziert werden kann, was beispielsweise bei Amplicon Librarys der Fall ist (doppelsträngiges PCR-Produkt), so muss dies bei der Konzentrationsumrechnung berücksichtigt werden (die Konzentration [Moleküle/ μl] wird halbiert).

Zur Bestimmung der idealen cpb-Zahl wird von Roche eine Titration vorgeschlagen. Das bedeutet, dass in kleinen Volumina emPCRs mit anschließendem Beadrecovery durchgeführt werden, wobei die Library in den unterschiedlichen Ansätzen mit verschiedenen cpb-Zahlen eingesetzt wird. Am Ende des Beadrecovery soll man über die Enrichment-Rate Rückschlüsse auf die Qualität der emPCR ziehen können –übliche Titrationsansätze beinhalten cpb Zahlen von 0,5; 1; 2; 4 oder ähnlichem [85]. Die cpb-Zahl, mit der die besten Ergebnisse im Beadrecovery erzielt werden, wird dann für einen großen Ansatz mit späterem Laufstart verwendet.

1.5 Weitere NGS-Methoden

Um eine Übersicht über andere derzeit auf dem Markt erhältlichen NGS-Methoden zu geben, folgt in den nächsten Kapiteln eine Vorstellung weiterer Systeme, wobei der Fokus auf das prinzipielle Vorgehen bei den unterschiedlichen Protokollen gelegt wird.

1.5.1 Helicos Bioscience

Das HeliScope Genetic Analysis System ist der erste Single-Molecule-Sequencer, der auf dem Markt erhältlich war. Das System basiert auf Arbeiten von Quake [95] und verwendet Cyclic Reversible Termination (CRT). Eine Besonderheit stellt die Tatsache dar, dass die Template-DNA für die Sequenzierung nicht klonal amplifiziert werden muss, ein möglicher Amplifikation-Bias wird somit vermieden. Abb. 32 zeigt das Prinzip der Sequenzierung mit dem HeliScope System.

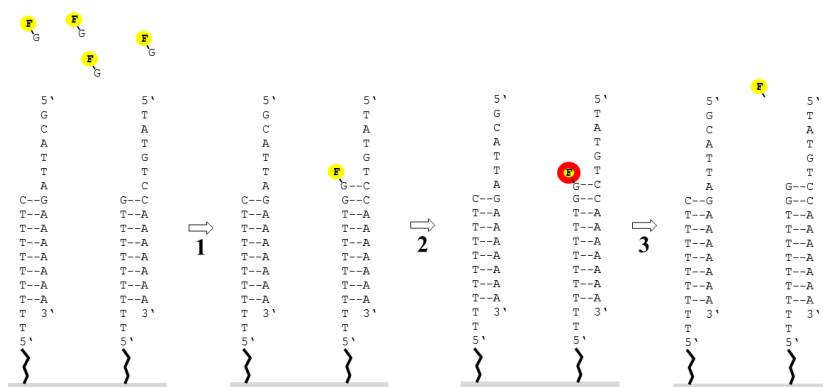


Abb. 32: Prinzip der Sequenzierung mit dem HeliScope System

Fragmentierte denaturierte genomische DNA, an die von einer terminalen Transferase ein als Adapter fungierender Poly-A-Schwanz angehängen wurde, hybridisiert an einen immobilisierten Poly-T-Primer. Der Primer ist über Streptavidin und Biotin an eine Glasoberfläche gebunden. Nun wird ein modifiziertes dNTP mit beispielsweise Guanin als Base dazugegeben und von einer Polymerase eingebaut, sofern C vorhanden ist (1). Dieses modifizierte dNTP trägt eine Fluoreszenzgruppe (F), die gleichzeitig an einen Terminator gekoppelt ist, der vorerst den weiteren Einbau von anderen dNTPs verhindert. Nach einem Waschschrift wird nun über Fluoreszenz die Position der immobilisierten DNA in der Flow-Cell und die eingebaute Base identifiziert (2). Mittels chemischer Spaltung wird nun die Fluoreszenzgruppe entfernt, die auch gleichzeitig den Terminator trägt (3), so dass der Einbau von weiteren neuen modifizierten dNTPs möglich ist (CRT). Jetzt können diese Schritte (1-3) mit einer anderen Base wiederholt werden, so dass nach und nach durch Strangaufbau die Sequenz bestimmt werden kann [96].

Einige hundert Zyklen (Abb. 32) führen zu Leselängen von durchschnittlich 35 bp mit einem Output von bis zu 35 Gb pro Run [97]. Mittels einer sogenannten Two-Pass-Strategy kann die Qualität der Sequenzen deutlich verbessert werden, es findet eine neue, reverse Sequenzierung der DNA statt: die zu sequenzierende DNA trägt hierbei zwei Adaptersequenzen; nach dem wie in Abb. 32 beschrieben der eine Strang sequenziert und damit aufgebaut wurde (Forward-Sequenzierung), wird die nun doppelsträngige DNA denaturiert und gewaschen, so dass nur der immobilisierte Strang übrigbleibt, welcher nach Hybridisierung eines Primers sequenziert werden kann (Reverse-Sequenzierung). Über die Position der Fluoreszenzsignale lassen sich die zugehörigen Forward- und Reverse-Sequenzen zuordnen.

1.5.2 Illumina

Der Illumina Genome Analyzer, auch unter dem Namen Solexa bekannt, kam 2006 auf den Markt, basiert auf Arbeiten von Turcatti [98,99] und der Fusion von vier verschiedenen Firmen (Solexa, Lynx Therapeutics, Manteia Predictive Medicine und Illumina). Zentrale Bestandteile sind die Amplifikation der zu sequenzierenden DNA mittels „Bridge-PCR“ [100,98] und die Verwendung von CRT, wobei im

Gegensatz zum HelicoScope-System die verschiedenen dNTPs bei Illumina unterschiedliche Fluoreszenzfarbstoffe tragen. In Abb. 33 ist das Prinzip der Sequenzierung mit dem Solexa-System erläutert.

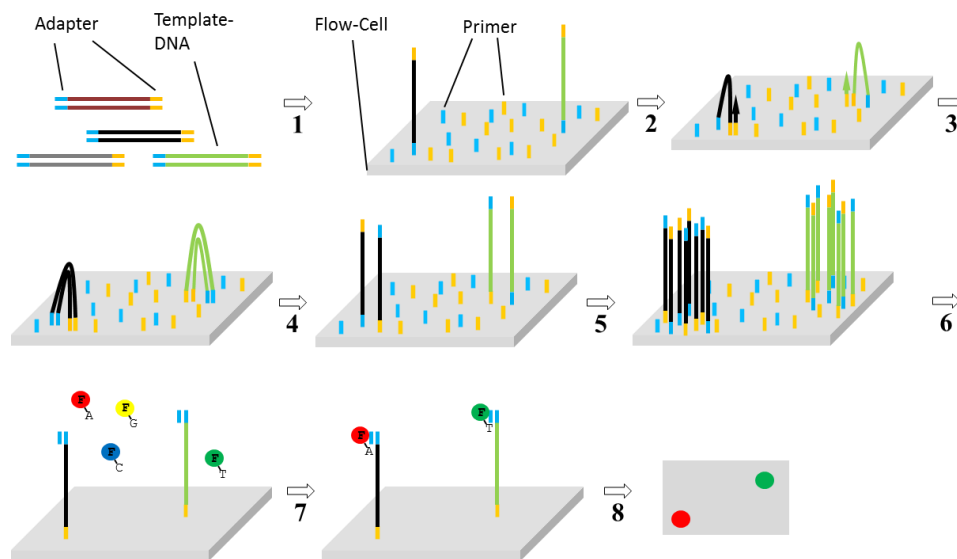


Abb. 33: Prinzip der Solexa-Sequenzierung

Eine Solexa-Library, bestehend aus Template-DNA (z.B. fragmentierte genomische DNA) und Adaptersequenzen (hellblau und gelb) wird auf einer Flow-Cell immobilisiert (1), was über modifizierte Enden der Adaptersequenzen geschieht. Mittels Bridge-PCR findet eine Amplifikation der Library statt: die einzelsträngige Library-DNA hybridisiert „brückenartig“ mit auf der Flow-Cell immobilisierten Primersequenzen (2), der Gegenstrang wird aufgebaut (3); nach Denaturierung (4) kommt es zu einer Wiederholung der entsprechenden Zyklen in der PCR (2-4), so dass man am Ende ein Cluster mit ca. 1000 klonal amplifizierten Librarysträngen erhält (5). Mehrere Millionen dieser lokal begrenzten Cluster können so innerhalb einer von 8 unabhängigen „Lanes“ entstehen. Zur Vereinfachung ist in den folgenden Schritten nur noch ein DNA-Strang jedes Clusters dargestellt. Zur Sequenzierung werden nach Hybridisierung eines entsprechenden Sequenzierungs-Primers die 4 unterschiedlichen modifizierten dNTPs gegeben (6); die dNTPs tragen entsprechende unterschiedliche Fluoreszenzfarbstoffe und an ihren 3'-Enden eine blockierende Gruppe. Es erfolgt der Einbau der dNTPs durch eine Polymerase (7); da die dNTPs eine blockierende Gruppe besitzen, kann nur ein dNTP eingebaut werden. Durch Fluoreszenz können nun die entsprechenden dNTPs, die in den Clustern eingebaut wurden, identifiziert werden (8). Nach Abspaltung von Fluoreszenzfarbstoffen und den blockierenden Gruppen können nun erneut Schritte der Sequenzierung erfolgen (6-8).

Es können Leseweiten bis zu 150 bp erreicht werden (250 bp sind angekündigt), und es können bis zu 600 Gb Output in einem Run generiert werden [101]. Es stehen Shotgun-, Amplicon-, Mate-Pair- und Paired-End-Protokolle zur Verfügung. Die größten Fehler, die bei der Sequenzierung auftreten, sind Substitutionen, besonders nach Einbau eines Guanins; außerdem können AT- und GC-reiche Regionen vermutlich aufgrund eines Amplifikations-Bias unterrepräsentiert sein [102-104].

1.5.3 Life/APG

Das von Life/APG entwickelte SOLID-Sequenziersystem (Support Oligonucleotide Ligation Detection) beruht auf Arbeiten von Shendure [105] und McKernan [106]. Es zeichnet sich durch klonale Amplifikation mittels emPCR an paramagnetischen Beads, welche nach Aufreinigung auf einem Glas-

Slide immobilisiert werden, und durch die Verwendung einer Ligase für das Sequenzieren anstatt einer Polymerase aus, was den Vorteil hat, dass auf dem Template-Strang in beide Richtungen sequenziert werden kann. Das Sequenzieren findet nach der sogenannten „Two Base Encoding“-Methode statt [106]; dabei werden 16 verschiedene Dinukleotid-Kombinationen über ein System von nur vier verschiedenen Fluoreszenzfarbstoffen bestimmt. Abb. 34 zeigt das Prinzip der Sequenzierung mit dem SOLID.

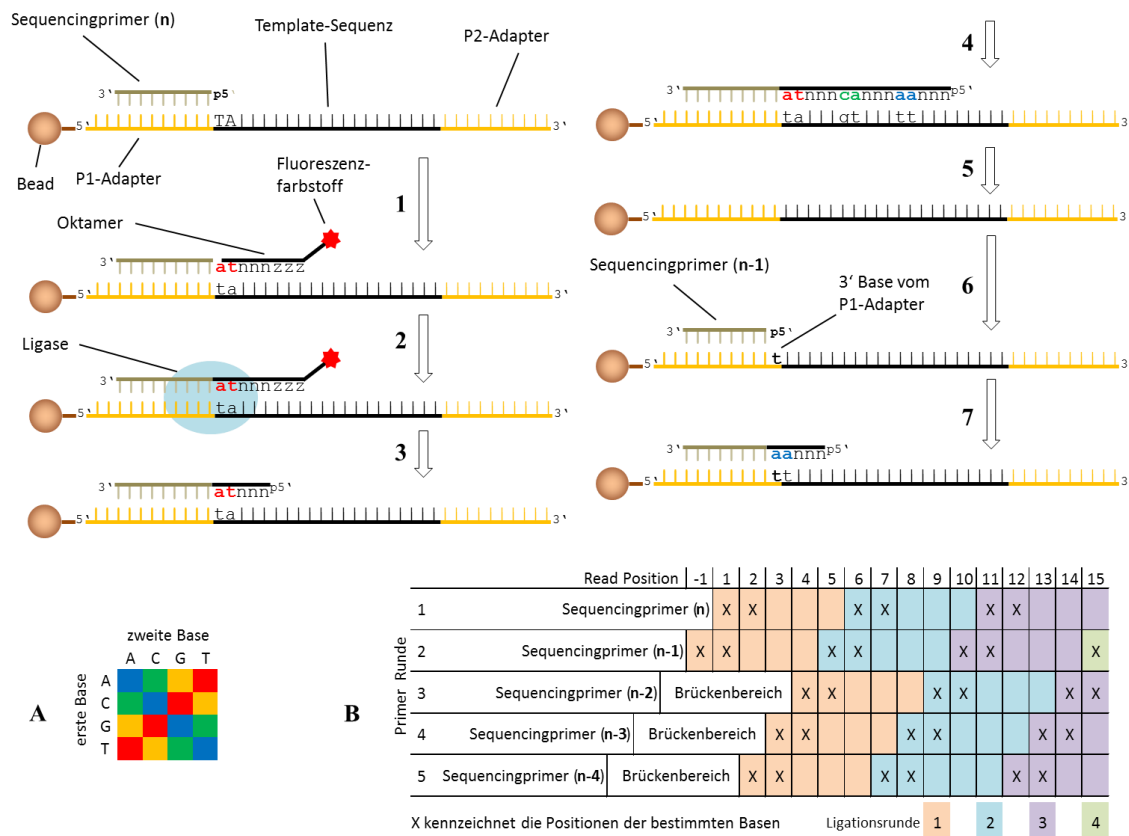


Abb. 34: Prinzip der SOLID-Sequenzierung

Eine SOLID-Library besteht aus Template-DNA, welche von zwei bestimmten Adaptersequenzen (P1 und P2) flankiert wird. Ein Sequencingprimer (n) hybridisiert an die Adapter-P1-Sequenz. Es startet die erste Ligationenrunde, wobei ein Oktamer, das an erster und zweiter Position die zum Template passenden Basen besitzt, mit der Template-DNA hybridisiert (1). Der Sequenz von erster und zweiter Base wird ein Fluoreszenzfarbstoff zugeordnet (A); die anderen 6 Nukleotide stellen universelle Basen dar, um die Komplexität zu verringern. Eine Ligase verbindet nun das Oktamer mit den vorhergehenden Basen (2). Jetzt wird über Fluoreszenz die Farbe bestimmt; in diesem Beispiel ist es rot. Das letzte „n“- und erste „z“-Nukleotid in 5'-Richtung sind über eine Phosphorothioat-Bindung miteinander verknüpft; diese wird mit Silbernitrat gespalten (3). Es beginnen erneute Ligationenrunden (insgesamt 15), es werden wie beschrieben die Oktamere eingebaut, deren erste beiden Basen sich über den Farbcodierung identifizieren lassen (4). Ist nach 15 Ligationenrunden die erste Primer-Runde beendet, so wird das Template denaturiert (5) und ein neuer Primer (n-1) hybridisiert (6). Dieser Primer zeichnet sich dadurch aus, dass er genau vor der letzten Base in 3'-Richtung von Adapter P1 mit diesem hybridisiert. Es folgen erneut Ligationenrunden, wobei die erste hierbei von besonderem Interesse ist (7): anhand des Farbcodes (A) können die ersten beiden Basen des Oktamers eindeutig identifiziert werden. In dem Beispiel wird in der ersten Ligationenrunde von Primer (n-1) ein blaues Signal detektiert. Laut Farbcodierung kann das theoretisch ein „aa“, „cc“, „gg“ oder „tt“ sein. Da die erste Base des Oktamers aber die letzte und bekannte des P1-Adapters ist, ein „t“, muss es sich bei der Oktamerkombination um ein „tt“ auf der Templateseite handeln, d.h. die erste Base im Template (Read) ist ein „t“. Durch immer weitere Ligationen- und Primer-Runden werden so alle Readpositionen doppelt bestimmt (B), so dass der gesamte Read sequenziert werden kann. In dem Beispiel gab die erste Ligationenrunde von Primer (n) ein rotes Signal. Da wie beschrieben die erste Readposition ein „t“ ist, muss also aufgrund des roten Signals die zweite Readposition (auf dem Template) ein „a“ sein (A).

Mit dem SOLID sollen durchschnittliche Leseweiten von ca. 75 bp [107] erreicht werden können; wie beim Solexa gibt es auch hier neben einem Shotgun- auch Paired-End- und Mate-Pair-Protokolle. Die größten Fehler hier sind auch wie beim Solexa Substitutionen, und eine mögliche Unterrepräsentation von AT- und GC-reichen Regionen [104].

1.5.4 Pacific Bioscience

Das von Pacific Bioscience entwickelte PACBIO RS-Sequenziersystem arbeitet nach der „Single Molecule Realtime“-Methode (SMRT). Das bedeutet, dass der Einbau der unterschiedlichen Nukleotide an einem einzigen DNA-Strang in Echtzeit aufgezeichnet wird; es ist also keine klonale Amplifikation nötig. Die Sequenzierung findet in SMRT-Zellen statt: eine DNA-Polymerase sitzt auf der Oberfläche eines sogenannten „Zero-Mode Waveguide Detector“ (ZMW); dieser ZMW stellt ein wenige Nanometer breites Loch in einem auf einer Glasplatte aufgetragenen Metallfilm dar [108]. Diese geringe Größe verhindert, dass sichtbares Laserlicht mit 600 nm Wellenlänge das ZMW passieren kann –es klingt exponentiell ab. Pacific Bioscience verwendet sogenannte „Real-Time Nucleotides“; diese tragen ihrer Base entsprechend unterschiedliche Fluoreszenzfarbstoffe, die über die Phosphatgruppen mit dem Nukleotid verbunden sind. Sie werden bei erfolgreichem Nukleotideinbau in die DNA einfach freigesetzt und stören damit nicht den weiteren Strangaufbau [109]. In Abb. 35 ist das Prinzip der SMRT-Methode von Pacific Bioscience erläutert.

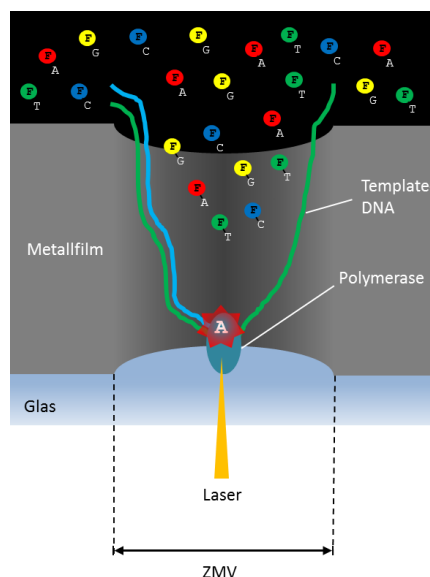


Abb. 35: SMRT-Methode von Pacific Bioscience

Das ZMW wird nur ca. 30 nm weit von dem Laserlicht durchdrungen. Dementsprechend findet auch nur Fluoreszenz bei den Nukleotiden statt, die sich innerhalb dieses Bereiches aufhalten; die übrigen, die ein Array von ZMWs fluten, werden nicht erkannt. Die Diffusion der dNTPs in das ZMW zu der Polymerase am Glasboden hin und zurück dauert wenige Microsekunden. Wird ein Nukleotid von der Polymerase in den DNA-Strang eingebaut, so verweilt es während des Nukleotideinbaus einige Millisekunden am Boden. Das ist im Gegensatz zu der Microsekunden dauernden Diffusion lang genug, um eine wesentlich höhere Signalintensität an Fluoreszenz zu erzeugen, wodurch der Einbau eines Nukleotids gegenüber der Diffusion im Hintergrund identifiziert wird.

Eine SMRT Zelle besteht aus einem Array von 75000 ZMWs, d.h. es können bis zu 75000 DNAs parallel sequenziert werden. Die durchschnittliche Leselänge wird mit knapp unter 1000 bp angegeben, wobei auch Leselängen bis zu 10000 bp möglich sein sollen. Über eine Zirkularisierung des Templates kann die gleiche DNA mehrfach sequenziert werden, wodurch die Genauigkeit über die Bildung einer Konsensus-Sequenz erhöht werden soll.

1.5.5 Ion Torrent

Bei den zuvor vorgestellten NGS-Technologien findet die Bestimmung der Sequenzierreaktion über die Detektion von Licht statt, es werden markierte (Oligo-) Nukleotide oder Enzymkaskaden zur Signalgeneration benötigt. Bei dem von Life Technologies entwickelten Ion Torrent-Sequencer wird der Nukleotideinbau über die Messung des pH-Wertes bestimmt [110]. An die zu sequenzierende DNA werden nach einer Fragmentation Adaptersequenzen ligiert, sie wird in einer emPCR an Acrylamidbeads amplifiziert und diese Beads in einem Enrichmentprozess über die Verwendung von magnetischen Beads angereichert. Die angereicherten Beads werden auf einen Chip aufgetragen, wo sie in Sensorwells sinken (bis zu 660 Mio. Wells auf dem proton II Chip). Während des Sequenzierens wird der Chip nacheinander mit allen vier Nukleotiden versorgt. Wird ein dNTP in die DNA von einer Polymerase eingebaut, so wird ein Proton freigesetzt. Das bewirkt eine Veränderung des pH-Wertes in dem Well, welche proportional zur Menge der eingebauten Nukleotide ist (Verringerung des pH-Wertes um 0,02 pro Nukleotid). Die Veränderung des pH-Wertes wird durch einen Sensor detektiert, der sich am Boden des Wells befindet. Auf diese Weise sollen auf einem Chip bis zu einer Gigabase (Gb) bestimmt werden können, mit Leseweiten der Reads von 200 bp Länge [www.iontorrent.com].

1.6 Vergleich der Sequencing-Kosten

Um einen Eindruck über die entstandenen Kosten bei der Sequenzierung von menschlichen Genomen zu bekommen und die unterschiedlichen Sequenziermethoden zu vergleichen, folgt eine tabellarische Auflistung von humanen Sequenzier-Projekten der letzten Jahre (Tab. 3).

Genom	Plattform	Template Librarys	Read- Anzahl [Mio]	Read- länge [Basen]	Basen- Coverage [x fach]	Assembly	Genom- Coverage [%]	Anzahl an Runs	Geschätzte Kosten [US\$]
J. Craig Venter	Automatisiertes Sanger	Matepairs von BACs, Fosmiden, Plasmiden	31,9	800	7,5	<i>De novo</i>	N/A	> 340000	70.000.000
James D. Watson	Roche/454	Fragmente	93,2	250	7,4	Aligned ¹⁾	95	234	1.000.000 ²⁾
männl. Yoruban	Illumina	93% Matepairs: 200 bp 7% Matepairs: 1,8 kb	3410 271	35 35	40,6	Aligned ¹⁾	99,9	40	250.000 ²⁾
männl. Han Chinese	Illumina	66% Fragmente 34% Matepairs: 135 bp & 440 bp	1921 1029	35 35	36	Aligned ¹⁾	99,9	35	500.000 ²⁾
männl. Koreaner	Illumina	21% Fragmente 79% Matepairs: 130 bp, 390 bp & 2,7 kb	393 1156	36 36, 88 & 106	27,8	Aligned ¹⁾	99,8	30	200.000 ²⁾
männl. Koreaner	Illumina	Matepair: 100 bp, 200 bp & 300 bp	1647	35,74	29	Aligned ¹⁾	99,9	15	250.000 ²⁾
männl. Yoruban	Life/APG	9% Fragmente 91% Matepairs: 600-3500 bp	211 2075	50 25, 50	17,9	Aligned ¹⁾	98,6	9,5	60.000 ²⁾
Stephen R. Quake	Helicos Bioscience	Fragmente	2725	32	28	Aligned ¹⁾	90	4	48.000 ²⁾
Frau mit akuter myeloischer Leukämie	Illumina	Fragmente einer Tumorprobe Fragmente gesunder Zellen	2730 1081	32 35	32,7 13,9	Aligned ¹⁾	91 83	98 34	1.600.000
Mann mit akuter myeloischer Leukämie	Illumina	Matepairs einer Tumorprobe Matepairs gesunder Zellen	1620 1351	35 50	23,3 21,3	Aligned ¹⁾	98,5 97,4	16,5 13,1	500.000

Tab. 3: Sequenzierstatistiken von humanen Genomprojekten [111]

1) Aligned an der vom Humanen Genomprojekt veröffentlichten Sequenz. 2) Nur Materialkosten.

Tabelle 4 zeigt einen Vergleich der derzeitigen Kosten pro Gb für das Sequenzieren von Sanger, Solid, Illumina und 454 am Max Planck Genomzenter Köln, einschließlich benötigter Laufzeiten für die Sequenzierung eines Humangenoms.

ABI 3730XL 96 capillary system (Sanger)	Solid 5500 xl und Illumina HighSeq 2000	Roche 454 Titanium plus
600-1100 bp durchschnittl. Read- Länge	36-100 bp durchschnittl. Read- Länge	750 bp modale Read- Länge
ca. 1,4 Mb / Tag ca. 0,042 Gb / Monat	ca. 200 Gb pro Run ca. 800 Gb pro Monat	ca. 0,75 Gb pro Run* ca. 15 Gb pro Monat
580.000 € pro Gb	300 € pro Gb	12.000 € pro Gb
17,5 Mio € in 60 Jahren für ein Humangenom	9000 € in einem Tag für ein Humangenom [§]	360.000 € in 20 Tagen für ein Humangenom [§]

Tab. 4: Sequenzierkosten pro Gb (Kommunikation mit Richard Reinhardt)

* bezogen auf die modale Leselänge; § Stand Juni 2011

1.7 SPRI

Die „Solid Phase Reversible Immobilization“-Methode, kurz SPRI, wurde erstmalig 1994 veröffentlicht [112]. Sie dient dazu, DNA in hoher Qualität aufzureinigen, was u.a. besonders bei der Template-Präparation für das Sequenzieren notwendig ist. Bei SPRI verwendet man magnetische Beads, die mit Polymeren beschichtet sind (Carboxylgruppen als funktionelle Gruppe). Die Carboxylgruppen haben einen pK_s -Wert von 4,7, was bedeutet, dass sie bei neutralen pH-Werten negativ geladen sind; da DNA ebenfalls bei neutralen pH-Werten negativ geladen ist, wird DNA unter diesen Bedingungen von den Beads abgestoßen. Gibt man sogenannte „Crowding-Reagents“, wie Polyethylenglycol, Alkohol, Salze oder Kombinationen dieser zu, so bindet die DNA an die Beads; unter bestimmten Bedingungen können fast 100% der DNA gebunden werden. Nach Waschschritten der Beads, die zu einer Verringerung der Ausbeute führen können (in Abhängigkeit der verwendeten Waschlösung), kann dann fast die gesamte an die Beads gebundene DNA wieder eluiert werden (aufgrund des pK_s -Wertes der Beads bei Verwendung von neutralen, schwachkonzentrierten Elutionspuffern oder Wasser). Am Ende der Aufreinigung kann man je nach Bedingungen Ausbeuten von 80% der aufgereinigten DNA erhalten [113].

Ferner kann über die Einstellung der Versuchsbedingungen auch eine größenselektive Isolation von DNA erreicht werden. Das kann durch eine Anpassung von Konzentration oder Molekulargewicht der Crowding-Reagents passieren, oder durch eine Veränderung von Salzkonzentrationen und pH-Werten, also generell über eine Veränderung der Polarität der Lösung; größere DNA-Moleküle lassen sich einfacher als kleinere bei geringen Konzentrationen der Crowding-Reagents [114] aus der Lösung an die Beads treiben, wodurch eine Selektion ermöglicht wird.

1.8 Zielsetzung dieser Arbeit

Beim 454-Pyrosequencing können cDNAs entweder ungerichtet über Random-Ligation der 454-Adapter-Sequenzen oder gerichtet über die Erstellung einer Amplicon Library sequenziert werden. Eine gerichtete Sequenzierung ist dann unverzichtbar, wenn z.B. aufgrund einer bestimmten Fragestellung immer von demselben Punkt einer bekannten Sequenz aus zu Vergleichszwecken ein unbekannter oder variabler Bereich bestimmt werden soll, wenn problematische Sequenzbereiche wie homopolymere Regionen vorhanden sind, oder bei Diversitätsbestimmungen von klinisch oder umwelttechnisch relevanten Proben, bei Mutationsanalysen, dem Auffinden von seltenen Allelen in Populationsstudien oder generell zur Analyse von Gen-Expression durch „RNA-Deepsequencing“ über cDNA-Sequenzierungen [115]. Die Erstellung einer Amplicon Library zur gerichteten Sequenzierung kann auch dann notwendig werden, wenn alte Standard-FLX-Librarys auf der Titanium Version (nach-) sequenziert werden müssen, weil das Standard-FLX-System nicht mehr mit Kits unterstützt wird und die Erstellung einer neuen Library aufgrund von schwieriger Probengewinnung schlecht möglich ist.

Wie in Kapitel 1.4.4.4 beschrieben, werden bei der Amplicon Library-Herstellung die 454-Adapter-Sequenzen als lange Overhangs an die Template-DNA heranamplifiziert. Gelingt dies nicht oder nur unzureichend, ist eine nachfolgende Sequenzierung nicht möglich oder führt, wenn überhaupt, zu sehr schlechten, nicht zufrieden stellenden Ergebnissen und damit verbundenen hohen Kosten.

Im Rahmen dieser Arbeit soll eine Alternative zur Amplicon Library-Herstellung entwickelt werden, die es ermöglicht, auch aus Template-DNAs 454-taugliche Librarys zu erstellen, bei denen eine Amplicon Library-Herstellung versagt hat. Den möglichen Gründen für das Versagen soll nachgegangen werden. Ferner soll die zu entwickelnde alternative Methodik eine gerichtete Sequenzierung erlauben entsprechend den 454-Amplicon Librarys, und die Sequencing-Ausbeuten und –Ergebnisse nicht schlechter ausfallen. Die Kosten und der Arbeitsaufwand zur Herstellung der Librarys sollen so gering wie möglich gehalten werden.

2. Material und Methoden

2.1 Material

2.1.1 Verbrauchsmaterial

Bezeichnung	Hersteller	Bestellnummer
0,2 ml PCR Tubes	Applied Bioscience	N80 10580 (Microamp 8-tube-strip) N801-0535 (Microamp 8-Cap-Strip)
0,5 ml Tubes	Sarstedt	72.699
1,5 ml Tubes	Sarstedt	72.706
10 ml Spritze	BD	309110
5 ml Tubes	VWR	16465-262
50 ml Spritze	BD	301035
50 ml Tubes	Corning	430829
Fusselfreie Tücher (Precision Wipes)	Kimtech Science	75512
Safeseal-Tips Premium	Biozym	692139 (10µl), 692151 (20µl), 692069 (200µl), 692078 (100µl)
Spitzen	Brico Medical Supplies	BN 1615
Sterilfilter 0,45µm	Schleich&Schuell	10462100

2.1.2 Geräte und Laborausstattung

Bezeichnung	Produkt	Hersteller	Bestellnummer
Bioanalyser	Agilent 2100 Bioanalyser	Agilent	2100
Feinwaage	SBC32	Scaltec	60406462
Fluorometer	Qubit	Invitrogen	Q32857
Heizblock	DriBlock DB2	Techne	FDB02AD
Laborwaage	universal	Satorius	448008
Labrotator		MPI-Eigenbau	
Microplate Centrifuge	X-12	Beckman Coulter	392472
Minifuge	Micro-Centrifuge	Roth	
Partikelzähler	Z1	Beckman Coulter	6605698
Rotor, Carrier & Cover	SX4750 ARIES	Beckman Coulter	369704, 392806, 392807
Sequencer	GS FLX+ Instrument	Roche	06372279001
Thermocycler	PTC-225 Peltier Thermal Cycler	MJ Research	
Tischzentrifuge	ependorf centrifuge 5417C	Eppendorf	
Tissue-Lyser	TissueLyser II, TissueLyser tube rack assemblies	Qiagen	85300, 69982
Vortexer	Vortex-Genie	Bencher&Hobein AG	

2.1.3 Chemikalien und Enzyme

Bezeichnung	Hersteller	Bestellnummer
Agarose (Genagarose LE)	Genaxxon	M3044.1000
Ampure XP beads	Agencourt	A63880
ATP	Roche	10519987001
Betaine	Sigma	B0300-5VL
ddaH ₂ O	MPI für molekulare Genetik	
EDTA	Merck	324503-100GM
Eisessig	Merck	1.00063.2500
Ethanol	Merck	1.00983.2500
Ethidiumbromid	Roth	22181.1
Isopropanol (2-Propanol)	Merck	1.09634.2500
KCl	Merck	529552-1KG
MgCl ₂ · 6H ₂ O	Merck	442615-500GM
NaCl	Merck	1.06404.1000
NaOH	Merck	1.06498.1000
Natriumacetat	Merck	1.06268.1000
Natrium-Chlorit-Tabletten (Micropur MP1)	Katadyn	8014996
Proteinase K	Qiagen	19133
Salzsäure (rauchend)	Merck	1.00317.2500
Sfil, BSA (100x), 10x NEB Puffer 2 & 4	New England Biolabs	R0123S
T4 DNA Ligase, 10x Ligase Puffer	Roche	10481220001
Tris	Merck	1.08382.0500
Tween 20	Pharmacia Biotech	17.1316-01

2.1.4 Kit-Systeme

Bezeichnung	Hersteller	Bestellnummer
DNA 7500 Kit	Agilent	5067-1506
GS FLX Titanium emPCR Breaking Kit LV/MV	Roche	05233658001
GS FLX Titanium emPCR Filters SV	Roche	05233674001
GS FLX Titanium General Library Preparation Kit	Roche	05233747001
GS FLX Titanium LV emPCR Kit (Lib-L)	Roche	05618428001
GS FLX Titanium PicoTiterPlate Kit 70x75	Roche	05233682001
GS FLX Titanium Rapid Library MID Adaptors	Roche	05619211001
GS FLX Titanium Rapid Library Preparation Kit	Roche	05608228001
GS FLX Titanium Sequencing Kit XL+	Roche	06380565001
GS FLX Titanium Sequencing Kit XLR70	Roche	05233526001
High Sensitivity DNA Kit	Agilent	5067-4626
MinElute PCR Purification Kit	Qiagen	28004
Quant-it-dsDNA HS Array	Invitrogen	Q32851
RNA 6000 Pico Kit	Agilent	5067-1513

2.1.5 Oligonukleotide und DNA-Marker

Mit * gekennzeichnete Basen sind phosphothioat-modifiziert; BioTEG bedeutet Biotin-TEG und P steht für ein Phosphat. Alle Oligonukleotide sind von MWG eurofins.

Bezeichnung	Hersteller/Sequenz	Bestellnummer
DNA Molecular Weight Marker XIV	Roche	11721933001
dATP	Genaxxon	M3018.0020
dCTP	Genaxxon	M3019.0020
dGTP	Genaxxon	M3020.0020
dTTP	Genaxxon	M3021.0020
454PrimerA	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3'	
454PrimerB	5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAG-3'	
ADA-MID1-CDNA	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGACGAGTGCGTCCATTACGGCCGGGGG-3'	
ADA-MID2-CDNA	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGACGCTCGACACCATTACGGCCGGGGG-3'	
ADA-MID3-CDNA	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGAGACGCACTCCATTACGGCCGGGGG-3'	
ADA-MID4-CDNA	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCACTGTAGCCATTACGGCCGGGGG-3'	
ADA-MID5-CDNA	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGATCAGACACGCCATTACGGCCGGGGG-3'	
ADA-MID6-CDNA	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGATATCGCGAGCCATTACGGCCGGGGG-3'	
AdapterA-MID20-primelSfi	5'-P-CTGTAGTCGTCTGAGTCGGAGACACGCAGGGATGAGATGG-3'	
AdapterA-MID20-SfiIA	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGACGACTACAGTTA-3'	
AdapterA-MID21-primelSfi	5'-P-CTAGTCTACGCTGAGTCGGAGACACGCAGGGATGAGATGG-3'	
AdapterA-MID21-SfiIA	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTAGACTAGTTA-3'	
AdapterA-MID22-primelSfi	5'-P-CATACTCGTACTGAGTCGGAGACACGCAGGGATGAGATGG-3'	
AdapterA-MID22-SfiIA	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGTACGAGTATGTTA-3'	
AdapterA-MID23-primelSfi	5'-P-CACGAGAGTACTGAGTCGGAGACACGCAGGGATGAGATGG-3'	
AdapterA-MID23-SfiIA	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGTACTCTCGTGTTA-3'	
AdapterA-MID24-primelSfi	5'-P-CTCGTCTCTACTGAGTCGGAGACACGCAGGGATGAGATGG-3'	
AdapterA-MID24-SfiIA	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGAGACGAGTTA-3'	
AdapterA-MID25-primelSfi	5'-P-CGAGCGACGACTGAGTCGGAGACACGCAGGGATGAGATGG-3'	
AdapterA-MID25-SfiIA	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGTCGCTCGTTA-3'	
AdapterA-MID26-primelSfi	5'-P-ACGCGTATGTCTGAGTCGGAGACACGCAGGGATGAGATGG-3'	
AdapterA-MID26-SfiIA	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGACATACGCGTTTA-3'	
AdapterA-MID3-primelSfi	5'-P-GAGTGCGTCTCTGAGTCGGAGACACGCAGGGATGAGATGG-3'	
AdapterA-MID3-SfiIA	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGAGACGCACTCTTA-3'	
AdapterB-primelSfi	5'-P-CTGAGACTGCCAAGGCACACAGGGGATAGG-3'	
AdapterB-SfiIB	5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAGAGG-3'	
ADB-CDNA	5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAGGAGGCGGCCTTTTTTTTTT-3'	
ADB-TCAG-TANKER	5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAGTTTTTTTTTTTTTTTTTTTTVN-3'	
rapidAdapterA- MID23-SfiIA	5'-CCATCTCATCCCTGCGTGTCTCCGACGACTTACTCTCGTG-TTA-3'	
rapidAdapterA-MID23-primelSfi	5'-P-CACGAGAGTAAGTCGTCTGGAGACACGCAGGGATGAGATGG-3'	
rapidAdapterA-MID24-primelSfi	5'-P-CTCGTCTCTAAGTCGTCTGGAGACACGCAGGGATGAGATGG-3'	
rapidAdapterA-MID24-SfiIA	5'-CCATCTCATCCCTGCGTGTCTCCGACGACTTAGAGACGAGTTA3'	
rapidAdapterA-MID25-primelSfi	5'-P-CGAGCGACGAAGTCGTCTGGAGACACGCAGGGATGAGATGG-3'	
rapidAdapterA-MID25-SfiIA	5'-CCATCTCATCCCTGCGTGTCTCCGACGACTTCGTCTCGTCTCGTTA-3'	
rapidAdapterA-MID26-	5'-P-ACGCGTATGTAGTCGTCTGGAGACACGCAGGGATGAGATGG-3'	

primelSfi	
rapidAdapterA-MID26-SfiIA	5'-CCATCTCATCCCTGCGTGTCTCCGACGACTACATACGCGTTTA-3'
rapidAdapterA-MID27-primelSfi	5'-P-ATACTCGCGTAGTCGTCGGAGACACGCAGGGATGAGATGG-3'
rapidAdapterA-MID27-SfiIA	5'-CCATCTCATCCCTGCGTGTCTCCGACGACTACGCGAGTATTTA-3'
rapidAdapterA-MID29-primelSfi	5'-P-ACTGTACAGTAGTCGTCGGAGACACGCAGGGATGAGATGG-3'
rapidAdapterA-MID29-SfiIA	5'-CCATCTCATCCCTGCGTGTCTCCGACGACTACTGTACAGTTTA-3'
rapidAdapterA-MID31-primelSfi	5'-P-AGACGACGCTAGTCGTCGGAGACACGCAGGGATGAGATGG-3'
rapidAdapterA-MID31-SfiIA	5'-CCATCTCATCCCTGCGTGTCTCCGACGACTAGCGTCGTCTTTA-3'
rapidAdapterA-MID32-primelSfi	5'-P-ATAGCGTACTAGTCGTCGGAGACACGCAGGGATGAGATGG-3'
rapidAdapterA-MID32-SfiIA	5'-CCATCTCATCCCTGCGTGTCTCCGACGACTAGTACGCTATTTA-3'
rapidAdapterA-MID33-primelSfi	5'-P-AGTACTCTATAGTCGTCGGAGACACGCAGGGATGAGATGG-3'
rapidAdapterA-MID33-SfiIA	5'-CCATCTCATCCCTGCGTGTCTCCGACGACTATAGAGTACTTTA-3'
rapidAdapterB- primelSfi	5'-P-AGTCGACTGCCAAGGCACACAGGGGATAGG-3'
rapidAdapterB-SfiIB	5'-CCTATCCCCTGTGTGCCTTGGCAGTCGACTAGG-3'
SfiIA-OConFor	5'-GATGGCCATTACGGCCCTCTGGACCTTGGCTGTCACTCAGTT-3'
SfiIB-OConRev	5'-GGTGGCCGAGGCGGCCCTTGGCTGTCACTCACTGCG-3'
Shortfor	5'-ATTACGGCCGGGGG-3'
Shortrev	5'-GCCGAGGCGGCCTT-3'
Tmid1A	5'-C*C*A*T*CTCATCCCTGCGTGTCTCCGACTCAGACGAGT*G*C*G*T-3'
Tmid1Aprime	5'-A*C*G*C*ACTCGTCTGAGTCG*G*A*G*A-3'
Tmid2A	5'-C*C*A*T*CTCATCCCTGCGTGTCTCCGACTCAGACGCTC*G*A*C*A-3'
Tmid2Aprime	5'-T*G*T*C*GAGCGTCTGAGTCG*G*A*G*A-3'
Tmid3A	5'-C*C*A*T*CTCATCCCTGCGTGTCTCCGACTCAGAGACGC*A*C*T*C-3'
Tmid3Aprime	5'-G*A*G*T*GCGTCTCTGAGTCG*G*A*G*A-3'
Tmid4A	5'-C*C*A*T*CTCATCCCTGCGTGTCTCCGACTCAGAGCACT*G*T*A*G-3'
Tmid4Aprime	5'-C*T*A*C*AGTGCTCTGAGTCG*G*A*G*A-3'
TprimerB	5'-5BioTEG-C*C*T*A*TCCCCTGTGTGCCTTGGCAGTC*T*C*A*G-3'
TprimerBprime	5'-C*T*G*A*GACT*G*C*C*A- 3'

2.1.6 Verwendete Puffer und Lösungen

Wenn nicht anders aufgeführt, wurden die unter 2.1.3 beschriebenen Chemikalien und Enzyme zur Herstellung der folgenden Puffer und Lösungen verwendet.

10 N NaOH

40 g NaOH-Pellets wurden auf einem elektrischen Rührer in 100 ml ddaH₂O gelöst und anschließend zur Aufbewahrung in eine Plastikflasche überführt.

EB-Puffer

Besteht aus 10 mM Tris-HCl pH 8.5. Für einen Liter wurden 1,211 g Tris-Base abgewogen, in 800 ml doppelt destilliertem Wasser gelöst, der pH-Wert mit rauchender Salzsäure eingestellt, auf das

Gesamtvolumen von 1 l aufgefüllt und die benötigten Mengen steril filtriert (mittels 50 ml Spritze und Sterilfilter).

1 x TE-Puffer

Besteht aus 10 mM Tris-HCl pH 8, 1 mM EDTA pH 8.

5 M NaCl-Lösung

14,61 g NaCl wurden in 50 ml ddaH₂O gelöst und anschließend steril filtriert.

3 M Natriumacetat-Lösung

123 g Natriumacetat wurden in 300 ml doppelt destilliertem Wasser gelöst, der pH-Wert von 5,2 mit Eisessig eingestellt, auf 500 ml Gesamtvolumen aufgefüllt und die benötigten Mengen steril filtriert.

5 mM ATP-Lösung

151,3 mg ATP wurden in 50 ml 25mM Tris-HCl pH 8 gelöst, steril filtriert, aliquotiert und bei -20°C eingefroren.

2.2 Methoden

2.2.1 emPCR und Beadrecovery

Alle verwendeten Chemikalien, Enzyme, Puffer und Lösungen stammten aus dem GS FLX Titanium LV emPCR Kit (Lib-L) und GS FLX Titanium emPCR Breaking Kit LV/MV, sofern sie nicht unter Kapitel 2.1.3 bzw. 2.1.6 aufgelistet sind; die Abläufe entsprechen den Roche-Protokollen [74].

2.2.1.1 Vorbereitung

Die bei -20°C gelagerten Enzyme und Reagenzien wurden bei Raumtemperatur aufgetaut, danach für 5 s gevortext und kurz herunterzentrifugiert (mit Ausnahme der Enzyme); die Enzyme wurden anschließend auf Eis gelagert.

Das Emulsions-Öl stand vor Gebrauch für eine Stunde bei Raumtemperatur.

2.2.1.2 Pre-Emulsion

Das Emulsions-Öl wurde mit den entsprechenden Adaptern in dem TissueLyser platziert und dort für 2 min bei 28 Hz geschüttelt. 2 ml 5x Mock-Amplification Mix wurden mit 8 ml ddaH₂O verdünnt, gevortext, und davon je 5 ml auf beide Emulsions-Öl-cups verteilt. Diese wurden ein paarmal invertiert und dann im TissueLyser für 5 min bei 28 Hz geschüttelt.

2.2.1.3 Live-Amplification Mix

Für zwei Emulsions-Öl-Cups wurden 2400 µl ddaH₂O, 3000 µl emPCR-Additive, 1560 µl 5x Amplification Mix, 460 µl Amplification Primer, 400 µl emPCR Enzyme Mix und 10 µl PPIase vermischt, für 5 s gevortext und anschließend auf Eis gestellt.

2.2.1.4 Library Capture

1 ml 10x Capture Bead Wash Buffer TW wurden mit 9 ml ddaH₂O zu 1x Capture Bead Wash Buffer TW verdünnt und gevortext.

Die DNA-Capture-Bead-Tubes wurden gevortext, die Beads in einer Minifuge für 10 s herunterzentrifugiert, um 180° gedreht und erneut herunterzentrifugiert. Der Überstand wurde vorsichtig abgezogen, ohne das Bead-Pellet zu zerstäuben. Jeder DNA-Capture Bead-Tube wurde zweimal mit 1 ml 1x Capture Bead Wash Buffer TW gewaschen (Wash Buffer dazugegeben, gevortext, herunterzentrifugiert und der Überstand abgezogen).

In jeden so vorbereiteten DNA-Capture-Bead-Tube wurde eine vorher berechnete entsprechende Menge 454-Library gegeben und die Tubes gevortext.

Das Library-Bead-Gemisch aus beiden Capture-Bead-Tubes wurde in je ein 5 ml Tube übertragen, und in jedes 5 ml Tube wurden jeweils 3,75 ml Live-Amplification Mix (2.2.1.3) gegeben. Dafür wurden zuerst jeweils 750 µl Live-Amplification Mix in die Capture-Bead-Tubes, die das Library-Bead-Gemisch enthielten, übertragen, gevortext und je in das 5 ml Tube übertragen, um restliche, an den Wänden klebende Bead- und DNA-Reste nicht zu verlieren –die Capture-Beads sind als weißer Film sichtbar. Dieser Schritt wurde für beide Capture-Bead-Tubes mit je 1 ml zweimal wiederholt (ergibt die genannten 3,75 ml Live-Amplification Mix je Tube); danach sollte kein weißer Film mehr an den Wänden sichtbar sein.

2.2.1.5 Emulsionsbildung

Der Inhalt von je einem zuvor präparierten 5 ml Tube (2.2.1.4) wurde in je ein Emulsions-Öl-Cup übertragen (2.2.1.2), ein paarmal invertiert und im TissueLyser für 5 min bei 12 Hz geschüttelt.

2.2.1.6 Amplifikation

Nach der Emulsionsbildung wurden die Emulsionen auf 0,2 ml PCR-Tubes verteilt zu je 100 µl pro Tube (oder Well) –ein Emulsions-Öl-Cup füllt ungefähr 180 Wells, also zwei 96er Platten. Das Verteilen erfolgte mit einer Multipipette. Danach wurden die 0,2 ml PCR-Tubes mit den entsprechenden Deckeln verschlossen, auf Blasen am Boden der Tubes überprüft (können durch leichtes Schlagen entfernt werden) und in einen Thermocycler gestellt. Folgendes Programm wurde verwendet: 1x (4 min bei 94°C); 50x (30 s bei 94°C; 4,5 min bei 58°C; 30 s bei 68°C); anschließend bei 10°C halten.

2.2.1.7 Emulsions-Kollektion

Die Emulsion wurde mittels Vakuumpumpe und den Aufsaugvorrichtungen aus den emPCR Breaking Kits in 50 ml Tubes überführt, wobei der Inhalt eines Emulsions-Öl-Cups auf je zwei 50 ml Tubes verteilt wurde (ein 50 ml Tube pro 96er Platte). Nach der Überführung wurden die 0,2 ml PCR-Tubes

bzw. 96er Platten zweimal mit je 100 µl Isopropanol pro Well gewaschen (Zugabe von Isopropanol mit einer Mehrfachpipette, Durchmischung durch zweimaliges Auf- und Abziehen, Überführung mittels Aufsaugvorrichtung in die entsprechenden 50 ml Tubes).

2.2.1.8 Waschen und Beadrecovery

Die 50 ml Tubes wurden gevortext, falls nötig mit Isopropanol auf insgesamt 40 ml aufgefüllt und die Capture-Beads bei 930 x g für 5 min in einer Tischzentrifuge pelletiert; anschließend wurde der Überstand vorsichtig entfernt. Es wurden pro 50 ml Tube je 35 ml Enhancing Fluid XT dazugeben, gevortext, bis das Pellet resuspendiert war und bei 930 x g für 5 min in einer Tischzentrifuge pelletiert. Nach dem Abziehen des Überstandes wurden 35 ml Isopropanol je Tube dazugegeben, das Pellet durch Vortexen resuspendiert, für 5 min bei 930 x g zentrifugiert; anschließend wurde der Überstand abgezogen. Nun wurden 35 ml Ethanol pro Tube dazugegeben, gevortext, bis das Pellet resuspendiert war und bei 930 x g für 5 min zentrifugiert. Der Überstand wurde abgenommen, 35 ml Enhancing Fluid XT zu den Beads dazugeben, gevortext und bei 930 x g für 5 min zentrifugiert. Der Überstand wurde abgenommen. Das DNA-Bead-Pellet wurde jetzt in 1,5 ml Tubes übertragen (ein 1,5 ml Tube pro 50 ml Tube): dafür wurde das Pellet in den 50 ml Tubes durch Vortexen resuspendiert, 1 ml in die 1,5 ml Tubes überführt, diese in einer Minifuge pelletiert und der Überstand verworfen (die Prozedur wurde solange wiederholt, bis das komplette Pellet eines 50 ml Tubes in ein 1,5 ml Tube übertragen wurde). Zum Schluss wurden die 1,5 ml Tubes in einer Minifuge pelletiert, und der Überstand verworfen.

2.2.1.9 Enrichment-Vorbereitung

Ein Heizblock wurde auf 65°C gestellt. Durch Mischen von 125 µl 10 N NaOH in 9,875 ml ddaH₂O wurde eine Melt-Lösung hergestellt.

Die DNA-Pellets enthaltenden 1,5 ml Tubes (2.2.1.8) wurden mit je 1 ml Melt-Lösung versetzt, gevortext und für zwei Minuten bei Raumtemperatur inkubiert. Anschließend wurden sie in einer Minifuge pelletiert und der Überstand mit einer Pipette abgezogen und verworfen. Es wurden wieder jeweils 1 ml Melt-Lösung dazugegeben und der Schritt wiederholt.

Zu den pelletierten Beads wurde jeweils 1 ml Annealing Buffer XT gegeben, gevortext, bis die Beads resuspendiert waren, in der Minifuge pelletiert und der Überstand verworfen. Es wurden wieder jeweils 1 ml Annealing Buffer XT dazugegeben, und die Beads durch Vortexen resuspendiert. 3 µl dieser Bead-Suspension wurden direkt nach dem Vortexen zum Bestimmen der Beadmenge mittels Partikelzähler verwendet; die Volumenbestimmung erfolgte über Wiegen, wobei ein Gramm mit einem Milliliter gleichgesetzt wurde (die Masse der Beads wird ignoriert). Der Quotient aus der Gesamtzahl der so gemessenen zurückgewonnenen Beads und den in der emPCR eingesetzten Beads ($3,5 \times 10^7$ pro Capture-Bead-Tube) ergibt die Beadrecovery Rate.

Die 1,5 ml Tubes wurden wieder pelletiert und der Überstand verworfen. Je Tube wurden 45 µl Annealing Buffer XT und 25 µl Enrichment-Primer dazugegeben, gevortext, die Tubes für 5 min in den 65°C warmen Heizblock gestellt und anschließend sofort auf Eis überführt, wo sie 2 min inkubierten. Danach wurden je 800 µl Enhancing Fluid XT in die Tubes gegeben, diese pelletiert und der

Überstand verworfen. Dann erfolgte ein zweimaliges Waschen mit je 1 ml Enhancing Fluid XT (Vortexen, Pelletieren und Überstand werfen). Es erfolgte erneut eine Zugabe von 800 µl Enhancing Fluid XT je 1,5 ml Tube, anschließend wurde kurz gevortext.

2.2.1.10 Vorbereitung der Enrichment-Beads

Das Enrichment-Bead-Tube wurde solange gevortext, bis die Enrichment-Beads vollständig resuspendiert waren. Anschließend wurden die Enrichment-Beads in einem MPC pelletiert; der Überstand wurde abpipettiert, wenn die Lösung klar war. Die Enrichment-Beads wurden zweimal im MPC mit 1 ml Enhancing Fluid XT gewaschen (nach Zugabe gevortext, im MPC pelletiert, Überstand abgezogen). Zu den pelletierten Enrichment-Beads wurden 320 µl Enhancing Fluid XT gegeben und gevortext.

2.2.1.11 Anreicherung der DNA-tragenden Capture-Beads

Zu jedem 1,5 ml Tube (2.2.1.9 Enrichment-Vorbereitung) wurden 80 µl der vorbereiteten Enrichment-Beads gegeben, gevortext bis die Suspension homogen war und die 1,5 ml Tubes in einen Labrotator für 5 min bei Raumtemperatur gestellt. Anschließend wurden die 1,5 ml Tubes in ein MPC gestellt und für 5 min pelletiert. Dann wurde mit einer Pipette der Überstand, der die nichtgebundenen weißen Capture-Beads enthielt, abgezogen, ohne dabei die an die Enrichment-Beads gebundenen Capture-Beads mitzunehmen.

Nun wurden die Beads mit je 1 ml Enhancing Fluid XT gewaschen; dazu wurden nach erfolgter Zugabe die 1,5 ml Tubes gevortext, im MPC für ungefähr ein bis zwei Minuten pelletiert (mit Invertieren) und dann wurde der Überstand vorsichtig abpipettiert. Diese Waschschrte wurden 6- bis 10-mal wiederholt, bis im Überstand keine weißen Beads mehr sichtbar waren.

2.2.1.12 Aufreinigung der angereicherten Beads

Die Tubes der gewaschenen Beads wurden mit je 700 µl Melt-Lösung versetzt, für mindestens 5 s gevortext und im MPC pelletiert, bis die Lösung nicht mehr bräunlich war. Der die Capture-Beads enthaltende Überstand wurde vorsichtig abgezogen und in ein neues 1,5 ml Tube überführt, wobei die Überstände der 1,5 ml Tubes der gewaschenen Beads, die aus einem Emulsion-Oil-Cup stammten, zusammengeführt wurden. Die zusammengeführten Überstände wurden pelletiert und die Überstände verworfen. Diese Prozedur wurde jetzt einmal wiederholt: die alten Tubes der gewaschenen Beads wurden erneut mit je 700 µl Melt-Lösung versetzt, für mindestens 5 s gevortext und im MPC pelletiert. Der Überstand wurde abgezogen und in die entsprechenden neuen 1,5 ml Tubes überführt, in denen sich die Beads des gerade zuvor erfolgten Melt-Schrittes befanden; die 1,5 ml Tubes wurden wieder in einer Minifuge pelletiert und die Überstände verworfen.

In jedes 1,5 ml Tube wurden 1 ml Annealing Buffer XT gegeben, gevortext, pelletiert und der Überstand abgezogen. Dieser Schritt wurde zweimal wiederholt. Danach wurden die Beads in je 200 µl Annealing Buffer XT durch Vortexen resuspendiert, je 50 µl Sequencing Primer dazugegeben, die Tubes für 5 min bei 65°C in einem Heizblock inkubiert und danach sofort für 2 min auf Eis getan. Anschließend wurden 800 µl Annealing Buffer XT je Tube dazugegeben, gevortext, in einer Minifuge

pelletiert und der Überstand verworfen. Die Beads wurden zweimal mit 1 ml Annealing Buffer XT gewaschen (Puffer dazugeben, vortexen, pelletieren, Überstand abziehen).

Als letztes wurden die Beads in 1 ml Annealing Buffer XT durch Vortexen vollständig resuspendiert; je Tube wurden 3 µl Beadsuspension im Partikelzähler zur Bead-Konzentrations- und Mengenbestimmung verwendet, das Volumen wurde wieder durch Wiegen bestimmt. Der Quotient aus der Gesamtzahl der so angereicherten Beads und der in Abschnitt 2.2.1.9 bestimmten Beads ergibt die „Enrichment Rate“.

2.2.2 Sequencing

Alle verwendeten Chemikalien, Enzyme, Puffer und Lösungen stammten aus dem GS FLX Titanium Sequencing Kit XLR70 und GS FLX Titanium PicoTiterPlate Kit 70x75, sofern sie nicht im Abschnitt 2.1.3 bzw. 2.1.6 aufgelistet sind; die Abläufe entsprechen den Roche-Protokollen [75].

2.2.2.1 Vorbereitung

Alle Reagenzien und Enzyme der Sequencing Reagents XLR70 wurden im Wasserbad bei Raumtemperatur aufgetaut und dann auf Eis gestellt. Der Titanium Bead Buffer musste für eine halbe Stunde auf Eis vorgekühlt werden. Danach wurden 1,2 ml Titanium Supplement CB und 34 µl Apyrase zu dem Titanium Bead Buffer gegeben, dieser ein paarmal vorsichtig invertiert und mit „BB2“ (Bead-Buffer 2) beschriftet.

2.2.2.2 Pre-Wash

Alte Sipper Tubes wurden gegen neue ausgetauscht; die Pre-Wash-Box wurde mit den entsprechenden Tubes vorbereitet, mit Pre-Wash-Buffer gefüllt und in den GS Sequencer gestellt. In der GS Sequencer Software wurde das Pre-Wash-Programm ausgewählt und gestartet.

2.2.2.3 Vorbereitung der PTP

Die PTP wurde in der mitgelieferten Packung mit BB2 übergossen, bis sie auf beiden Seiten luftblasenfrei von dem Puffer bedeckt war. Anschließend wurde sie mittels Vakuumpumpe entgast. Dort verblieb die PTP, bis sie beladen wurde.

Der Plastik-Spacer, das Cartridge Seal und die Bead-Deposition-Device (BDD) wurden mit ddaH₂O abgespült und zum Trocknen auf ein fusselfreies Tuch gelegt.

2.2.2.4 Vorbereitung der Packing-Beads

Die Packing-Beads wurden dreimal mit 1 ml BB2 durch Zentrifugation in einer Tischzentrifuge bei 9300 x g für 5 min gewaschen (BB2 dazugeben, vortexen, zentrifugieren, Überstand abgießen). Anschließend wurden sie in 550 µl BB2 resuspendiert.

2.2.2.5 Vorbereitung der DNA-Beads

Für eine zwei-Regionen-PTP wurden 2,2 Mio. DNA-Beads pro Region verwendet. Das benötigte Volumen an DNA-Bead-Suspension wurde anhand der Konzentrationen der jeweiligen Beads (Abschnitt 2.2.1.12) bestimmt. Für jede Region wurde ein entsprechendes 2 ml Tube bereitgestellt, mit der errechneten DNA-Bead-Menge befüllt und noch zusätzlich mit Control-DNA-Beads versehen (20 µl pro Region bei zwei Regionen und 10 µl pro Region bei vier Regionen). Die 2 ml Tubes wurden in einer Minifuge herunterzentrifugiert und der Überstand vorsichtig abgezogen bis am Ende eine Menge von 50-100 µl in den Tubes verblieb.

Aus 1570 µl BB2, 150 µl Polymerase Cofactor und 300 µl DNA Polymerase wurden in einem 2 ml Tube der „DNA Bead Incubation Mix“ (DBIM) hergestellt.

Zu jedem DNA-Bead-Tube wurden 950 µl DBIM gegeben (320 µl bei vier Regionen) und diese für mindestens 15 min in einem Labrotator bei Raumtemperatur inkubiert (nicht länger als 50 min).

Anschließend wurden in jedes DNA-Bead-Tube 265 µl Packing-Beads und 435 µl BB2 gegeben, und diese bis zur Beladung der PTP für mindestens 5 min auf einem Labrotator bei Raumtemperatur inkubiert.

2.2.2.6 Vorbereitung von Enzyme- und PPIase-Beads

Die Beads wurden gevortext, bis sie vollständig resuspendiert waren und wurden im MPC je dreimal mit 1 ml BB2 gewaschen (Beads in BB2 durch Vortexen resuspendiert, im MPC pelletiert, Überstand abgezogen). Nach dem letzten Waschschrift wurden die Enzyme-Beads in je 1 ml BB2 und die PPIase-Beads in 500 µl BB2 resuspendiert. Drei 5 ml Tubes wurden mit „Layer 1“, „Layer 3“ und „Layer 4“ beschriftet. Layer 1 wurde aus 3250 µl BB2 und 550 µl Enzyme-Beads hergestellt; Layer 3 aus 2500 µl BB2 und 1300 µl Enzyme-Beads, und Layer 4 aus 3340 µl BB2 und 460 µl PPIase-Beads.

2.2.2.7 Zusammenbau und Beladen der PTP

Das BDD und die Plastik-Spacer wurden vorsichtig mit einem fusselfreien Tuch abgetupft, die PTP aus dem BB2-Bad geholt und in das BDD mit dem Plastik-Spacer (zwei oder vier Regionen) eingebaut.

Nach Vortexen wurden 1860 µl der Layer 1-Bead-Suspension pro Region auf die PTP mittels Pipettieren aufgetragen, die Lade- und Belüftungslöcher der PTP mit Tesafilm abgedichtet, die BDD in die selbstaustarierenden Microplate-Zentrifuge gestellt und für 5 min bei 1620 x *g* zentrifugiert. Anschließend wurde der Tesafilm entfernt und die Überstände herauspipettiert. Nach Vortexen wurden 1700 µl der DNA-Bead-Suspension pro Region auf die PTP aufgetragen, die Lade- und Belüftungslöcher der PTP abgedichtet, die BDD in die Microplate-Zentrifuge gestellt und für 10 min bei 1620 x *g* zentrifugiert. Nach vorsichtigem Entfernen des Tesafilms wurden die Überstände herauspipettiert. Nach Vortexen wurden 1860 µl der Layer 3-Suspension pro Region auf die PTP aufgetragen, die Lade- und Belüftungslöcher der PTP abgedichtet, die BDD in die Microplate-Zentrifuge gestellt und für 5 min bei 1620 x *g* zentrifugiert. Anschließend wurde der Tesafilm entfernt und die Überstände herauspipettiert. Das selbe Vorgehen wie bei Layer 3 geschah nun mit Layer 4.

2.2.2.8 Vorbereitung des GS Sequencers und Runstart

Nach beendetem Pre-Wash wurde die Pre-Wash-Kassette entfernt. Jede Titanium Buffer CB-Flasche wurde mit 6,6 ml Titanium Supplement CB und 1 ml DTT versetzt und geschüttelt. Eine Natriumchlorit-Tablette wurde in dem „Post Run Wash“-Tube aufgelöst; 13,2 µl von vorher in „Inhibitor TW“-Puffer verdünnter PPIase (Verhältnis 1:5) wurden in das „Inhibitor TW“-Tube gegeben; in das „Buffer for Apyrase“-Tube wurden 260 µl Apyrase und in das „Buffer for dATP (A)“-Tube wurden 3 ml dATP pipettiert. Die Tubes in der Reagenzienhalterung wurden anschließend durch 20ig-faches Invertieren gemischt und zusammen mit den Titanium Buffer CB-Flaschen in die Sequencing-Kassette gestellt, welche in den GS Sequencer getan wurde.

Die Kamera-Tür wurde geöffnet, die alte PTP und das alte Cartridge Seal aus dem PTP-Rahmen entfernt und dieser sowie die Kamera-Frontplatte mit Ethanol (50%) unter Verwendung eines fusselfreien Tuches gereinigt.

Waren diese luftgetrocknet, wurde das neue Cartridge Seal eingesetzt, auf die Oberfläche des PTP-Rahmens eine Tween-20-Lösung (10%) aufgetragen, die beladene PTP eingesetzt und die PTP-Tür geschlossen.

In der GS Sequencer-Software wurden die entsprechenden Parameter eingegeben –Runname, Runtyp (Titanium), Anzahl an Regionen, Anzahl Flow-Cycles (Maximum: 200), Processing-Typ (Image Processing only), und der Run gestartet.

2.2.2.9 Signalprocessing und MID-Analyse

Das Signalprocessing fand auf den externen Servern des MPI für molekulare Genetik statt. Das Signalprocessing und das Aufteilen der Reads nach ihren MIDs wurde von Sven Klages (MPI für molekulare Genetik) durchgeführt; Sven Klages hatte auch das Perl-Script für die MID-Zerlegung geschrieben. Beim Signalprocessing wurden die von Roche vorgegebenen Standard-Parameter verwendet.

2.2.3 SPRI-Aufreinigung

Die von kleinen Fragmenten aufzureinigende DNA wurde in einem 1,5 ml Tube mit einer bestimmten Menge an SPRI-Beads (Ampurebeads) versetzt, wobei sich das Verhältnis von DNA-Menge zu Ampurebead-Menge aus einer vorher erfolgten Titration ergibt, die mit jeder Ampurebead-Charge durchgeführt werden musste (vergleiche Abschnitt 2.2.4). Das Gemisch wurde für 5 s gevortext, und dann für 5 min bei Raumtemperatur stehen gelassen. Anschließend wurde das 1,5 ml Tube im MPC pelletiert. War der Überstand klar, wurde er abgezogen und verworfen (im Überstand befinden sich die kleineren Fragmente). Die Ampurebeads wurden zweimal mit 70 %igem Ethanol gewaschen (Zugabe von Ethanol in das MPC, Warten von 20 s und Abzug des Ethanols) und anschließend für zwei bis fünf Minuten bei Raumtemperatur im MPC getrocknet. Die Ampurebeads wurden in EB-Puffer durch Vortexen vollständig resuspendiert und im MPC pelletiert (die Menge an EB-Puffer hing von den

später folgenden Versuchen ab und kann frei gewählt werden; sie sollte aber ein Volumen von 10 µl nicht unterschreiten). Der die DNA enthaltene Überstand wurde in ein neues 1,5 ml Tube pipettiert.

2.2.4 Ampurebead-Titration

Der Ablauf der Ampurebead-Titration entspricht den Vorgaben von Roche [85]. 48 µl von Roche's DNA Molecular Weight Marker XIV wurden mit 1152 µl ddaH₂O verdünnt und je 100 µl verdünnter Marker auf elf 1,5 µl Tubes aliquotiert. Nach Vortexen wurden in die Tubes mit dem aliquotierten Marker je 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 und 100 µl Ampurebeads gegeben. Nun wurde eine SPRI-Aufreinigung durchgeführt (Kapitel 2.2.3) und die DNA mit je 10 µl EB-Puffer eluiert. Von diesen elf Aufreinigungen wurden je 1 µl auf einen DNA 7500-Chip aufgetragen und analysiert; zusätzlich wurde noch 1 µl von verdünntem DNA Molecular Weight Marker XIV (4 µl Marker in 6 µl ddaH₂O) als Kontrolle bei der Analyse mit aufgetragen. Der „Height Threshold“ in der „Expert 2100“-Software wurde auf 5 gesetzt. Laut Roche soll die „Peak Ratio“, der Quotient der Marker-Konzentrationen, bei 200bp/900bp bei Null, bei 300bp/900bp bei 0,4, bei 400bp/900bp bei 0,7 und bei 500bp/900bp bei 2,9 für eine SPRI-Aufreinigung von 454-Librarys liegen. Die Ergebnisse einer Titration werden auf Seite 64 ff. gezeigt.

2.2.5 Herstellung von MID-Adapttern

Die Herstellung der 454-Adapter, die nach Adapter-A-Sequenz und Key ein 10 bp langes MID tragen, erfolgte nach den entsprechenden Vorgaben von Roche [83]. Auf die verwendeten Sequenzen wird in den jeweiligen Abschnitten verwiesen; die Sequenzen der 10 bp langen MIDs entsprachen den von Roche vorgegebenen [84].

10 ml Oligo-Puffer wurden aus 9,9 ml 1x TE-Puffer und 0,1 ml 5 M NaCl hergestellt. In ein 0,2 ml PCR-Tube wurden 60 µl Oligo-Puffer, 20 µl T-MID-A-Oligo (100 µM) und 20 µl T-MID-Aprime-Oligo (100 µM) gegeben. In ein weiteres 0,2 ml PCR-Tube wurden 60 µl Oligo-Puffer, 20 µl T-B-Oligo (100 µM) und 20 µl T-Bprime-Oligo (100 µM) gegeben. Beide 0,2 ml PCR-Tube wurden gevortext und in einem Thermocycler mit folgendem Programm inkubiert: 95°C für 1 min, -0,1°C/s bis auf 15°C, Halten bei 14°C. Der Inhalt der beiden PCR-Tubes wurde kombiniert und gevortext.

2.2.6 Sequenzierungsversuche des aqf-Projektes

Es sollten cDNAs für das aqf-Projekt („AquaFirst“) sequenziert werden, damit das NGS-Sequencing auf dem 454 mit Microarrays für das Auffinden von ESTs (expressed sequence tags) verglichen werden konnte (das Ziel des AquaFirst-Projektes ist die Identifizierung von Genen in Auster, Forelle, Seebrasse und Wolfsbarsch, die mit Stress- und Krankheitsresistenzen zusammenhängen [137]). Die cDNA wurde in der Abteilung Richard Reinhard von der Arbeitsgruppe Kube unter Verwendung des „MINT Universal cDNA synthesis kit“ von Evrogen (Cat#SK002) hergestellt. Die cDNAs besaßen eine Länge zwischen 2 bis 0,5 kb, da sie in einem Fraktionssammler größenselektiert wurden, und trugen die bei der cDNA-Herstellung über das MINT-Kit eingeführten, die Insert-DNA flankierenden Sequenzen an den 5'- und 3'-Enden. Der 5'-Anfang der cDNAs besaß folgende Sequenz:

5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGG-Insert-3'

Das 3'-Ende der cDNAs besaß folgende Sequenz, wobei „N“ jede Base sein kann und „B“ jede Base außer Adenin ist:

5'-Insert-NB(A)₂₀GGCCGCCTCGGCCACTCTGCGTTGATACCACTGCTT-3'

2.2.6.1 Erstellung von Titanium General Librarys

Von den aqf-cDNAs aqf3b, aqf4b, aqf5b und aqf6b wurden jeweils eine Titanium General Library hergestellt. Alle verwendeten Chemikalien und Enzyme stammten aus dem GS FLX Titanium General Library Preparation Kit, sofern sie nicht unter Kapitel 2.1.3 bzw. 2.1.6 aufgelistet sind; die 454-Library-Herstellung folgte dem entsprechenden Roche Protokoll [85].

Von jeder cDNA wurden je 5 µg in einem Volumen von 100 µl durch Ultraschall für 1 min fragmentiert, dann mit SPRI-Beads aufgereinigt (mit der durch die Titration bestimmten Menge Ampurebeads) und mit 24 µl EB-Puffer eluiert. 1 µl jeder cDNA wurden auf einen 7500 DNA-Chip aufgetragen und analysiert.

Mit den verbliebenen 23 µl cDNA wird ein Endpolishing durchgeführt; es wurden zur cDNA je 5 µl 10x Polishing Buffer, 5 µl BSA, 5 µl ATP, 2 µl dNTP Mix, 5 µl T4 PNK und 5 µl T4 DNA Polymerase dazugegeben, gevortext und diese im Thermocycler für 15 min bei 12°C und darauf folgend 15 min bei 25°C inkubiert. Anschließend erfolgte eine Aufreinigung der Ansätze mit dem MinElute PCR Purification Kit nach den Angaben des Herstellers, wobei jeweils 250 µl PBI-Puffer verwendet wurden und die Elution mit jeweils 10 µl EB-Puffer erfolgte.

Es folgte eine Adapter-Ligation mit selbst hergestellten MID-Adapttern. Zu den 10 µl cDNA wurden jeweils 20 µl 2x Ligations-Puffer, 5 µl MID-Adapter und 5 µl Ligase dazugegeben, gevortext und für 15 min im Thermocycler bei 25°C inkubiert. Die verwendeten Adapter trugen folgende MIDs: aqf3b bekam MID1, hergestellt aus den Oligonukleotiden TprimerB, TprimerBprime, Tmid1A und Tmid1Aprime; aqf4b bekam MID2, hergestellt aus den Oligonukleotiden TprimerB, TprimerBprime, Tmid2A und Tmid2Aprime; aqf5b bekam MID3, hergestellt aus den Oligonukleotiden TprimerB,

TprimerBprime, Tmid3A und Tmid3Aprime; aqf6b bekam MID4, hergestellt aus den Oligonukleotiden TprimerB, TprimerBprime, Tmid4A und Tmid4Aprime. Danach erfolgte eine Aufreinigung der Ligation-Ansätze mit dem MinElute PCR Purification Kit nach den Angaben des Herstellers, wobei jeweils 200 µl PBI-Puffer verwendet wurden und die Elution mit jeweils 100 µl EB-Puffer durchgeführt wurde.

Nach einer weiteren SPRI-Aufreinigung der vier cDNAs, wobei diese mit je 25 µl EB-Puffer eluiert wurden, wurden sie an Magnetic-Beads immobilisiert. Dazu wurden je 50 µl Library-Immobilisation-Beads in ein 1,5 ml Tube überführt und im MPC pelletiert. Nach Abzug des Überstandes erfolgte ein zweifaches Waschen der Beads mit je 100 µl 2x Library Binding Buffer (Puffer dazugeben, vortexen, im MPC pelletieren und Überstand abziehen). Die Beads wurden in je 25 µl 2x Library Binding Buffer resuspendiert. Zu den je 25 µl Beads wurden die 25 µl SPRI-aufgereinigte cDNAs gegeben und in einem Labrotator für 20 min bei Raumtemperatur inkubiert. Anschließend wurden die Beads im MPC zweimal mit je 100 µl Library Wash Buffer gewaschen (Zugabe des Puffers, Pelletieren im MPC und Abzug des Überstandes). Danach wurde der Überstand vollständig abgezogen, und in jedes Tube zu der an die Library-Immobilisation-Beads gebundenen cDNA-Librarys das Fill-In-Reaction-Mix gegeben, bestehend aus jeweils 40 µl Molecular Biology Grade Water, 5 µl 10x Fill-in Polymerase Buffer, 2 µl dNTP Mix und 3 µl Fill-in Polymerase. Nach Vortexen wurden die Tubes bei 37°C für 20 min inkubiert (Thermocycler oder Wasserbad). Anschließend wurden die Beads zweimal mit je 100 µl Library Wash Buffer gewaschen (Zugabe des Puffers, Pelletieren im MPC, Abzug des Überstandes); nach dem letzten Waschschrift wurde der Überstand so gut es geht vollständig abgezogen.

Es wurden der cDNA-Library-Anzahl entsprechend vier neue 1,5 ml Tubes mit sogenannter Neutralization-Solution vorbereitet; dazu wurde in jedes Tube zu 500 µl PBI-Puffer (aus dem MinElute PCR Purification Kit) 10 µl 3 M Natriumacetat pH 5,2 gegeben und gevortext.

In die 1,5 ml Tubes, die die an Beads immobilisierte Librarys enthielten, wurden jeweils 50 µl MELT-Lösung gegeben (125 µl 10 N NaOH und 9,875 ml ddaH₂O), die Tubes gut gevortext und im MPC pelletiert. Der 50 µl Überstand wurde in die neuen Tubes mit der Neutralization-Solution übertragen. Diese Schritte wurden einmal wiederholt, so dass die Tubes mit der Neutralization-Solution jeweils insgesamt 100 µl MELT-Überstand enthielten. Die MELT-gewaschenen Tubes, die die Beads enthielten, wurden verworfen.

Die so neutralisierten einzelsträngigen cDNA-Librarys wurden nun mit dem MinElute PCR Purification Kit nach den Angaben des Herstellers aufgereinigt, wobei auf die Zugabe von PBI-Puffer verzichtet wurde und die Elution der Librarys in 15 µl EB-Puffer stattfand. Zur Qualitätskontrolle und gleichzeitig als Konzentrationsbestimmung wurden die Librarys auf einen RNA Pico 6000 Chip aufgetragen und analysiert.

Die Ergebnisse der General Library-Produktion sind auf Seite 68 ff. dargestellt.

2.2.6.2 Titration der Titanium General Librarys

Mit jeder der hergestellten Titanium General Librarys aus den aqf-cDNAs wurde eine Titration bei 0,5, 1, 2 und 4 cpb durchgeführt (also vier Titrationsansätze pro Library).

Die Ansätze der emPCR und des Beadrecoverys folgten dem auf Seite 37 unter Abschnitt 2.2.1 beschriebenen Ablauf, mit den folgenden Änderungen, die für die Titration einer Library mit vier verschiedenen cpb-Zahlen beschrieben ist, da pro Ansatz nur 1/64 der Volumina und Mengen verwendet werden:

Bei 2.2.1.2 wurden nach dem Schütteln des Emulsionsöls je 312,5 µl Öl in vier 2 ml Tubes gegeben; es wurden 125 µl 5x Mock-Mix in 500 µl ddaH₂O verdünnt und davon je 156,2 µl zu den 2 ml Tubes gegeben, und diese in den entsprechenden Adapter im Tissue-Lyser wie angegeben geschüttelt. Bei 2.2.1.3 wurde die für die Titration benötigte Menge angesetzt, also für die Titration einer Library bei vier cpb-Zahlen 1/16 der angegebenen Volumina. Bei 2.2.1.4 wurden vier 1,5 ml Tubes mit je 36,56 µl Capture-Beads befüllt, was einer Menge von $1,09 \times 10^6$ Beads pro Emulsion darstellt. Diese wurden wie beschrieben mit 200 µl Wash Buffer gewaschen. Zu ihnen wurden die in Tab. 5 (Seite 47) angegebenen Mengen Library und jeweils 117 µl Live-Amplification Mix gegeben. Das Library-Bead-Gemisch der 1,5 ml Tubes wurde jeweils in die entsprechenden 2 ml Öl-Mock-Mix-Tubes überführt und in den entsprechenden Adaptern im Tissue-Lyser wie in Abschnitt 2.2.1.5 beschrieben geschüttelt. Bei 2.2.1.6 wurde das Verteilen in die PCR-Tubes mit einer Pipette durchgeführt.

Name	MID	Konzentra- tion in ng/µl	durchschnittl. Länge in bp	Konzentration in Moleküle/µl	µl an Library für			
					0,5 cpb	1 cpb	2 cpb	4 cpb
aqf3b	1	1,971	700	5164858796,40	1,06E-04	2,11E-04	4,22E-04	8,44E-04
aqf4b	2	1,325	700	3472063878,86	1,57E-04	3,14E-04	6,28E-04	1,26E-03
aqf5b	3	0,456	700	1194914059,44	4,56E-04	9,12E-04	1,82E-03	3,65E-03
aqf6b	4	0,069	700	180809364,26	3,01E-03	6,03E-03	1,21E-02	2,41E-02

Tab. 5: Berechnung der Titrationsansätze der General Libraries

Aus den mittels RNA-Chip bestimmten Konzentrationen und anhand der aus den Verläufen der Messkurven von RNA- und DNA-Chips abgeschätzten durchschnittlichen Längen der vier aqf-Libraries wurde die Konzentration in Molekülen pro Volumen errechnet (vergleiche Kapitel 1.4.5 auf Seite 24).

Emulsions-Kollektion (2.2.1.7), Waschen und Beadrecovery (2.2.1.8) erfolgten nicht durch Zentrifugieren, sondern durch das Waschen in einer Spritze mit Filteraufsatz. Der Inhalt von PCR-Tubes mit Emulsionen einer cpb-Zahl wurde mit einer Spritze aufgezogen, dann pro PCR-Tube je 100 µl Isopropanol gegeben und der Inhalt in die Spritze gezogen. Die Spritze wurde mit der Nadel nach oben gehalten, ca. 3 ml Luft eingezogen, ein Filter aus dem GS FLX Titanium emPCR Filters SV-Kit auf die Spritze gesetzt, die Spritze gevortext, der Inhalt der Spritze vorsichtig durch den Filter gedrückt und verworfen. Es wurden erneut je 100 µl Isopropanol in die PCR-Tubes gegeben. Der Filter wurde von der Spritze abgenommen und das Isopropanol aus den PCR-Tubes in die gleiche Spritze aufgezogen, die Spritze mit der Nadel nach oben gehalten, ca. 3 ml Luft eingezogen und danach der gleiche Filter aufgesetzt. Die Spritze wurde gevortext, der Inhalt vorsichtig durch den Filter gedrückt und verworfen. Es folgte ein dreimaliges Waschen mit Isopropanol (dafür wurden ca. 8 ml aufgezogen, gevortext und wieder abgegeben) und analog dazu einmal ein Waschen mit Ethanol und 1x Enhancing Fluid. Zum Resuspendieren der Beads aus dem Filter in ein neues 1,5 ml Tube wurde sehr wenig 1x Enhancing Fluid XT in die Spitze der Spritze eingezogen, dann sehr schnell und ruckartig Luft eingezogen, was zu einem Ablösen der Beads von dem Filter führte und der Inhalt der

Spritze wurde nach Abnehmen des Filters in ein neues 1,5 ml Tube gegeben. Diesen Vorgang wurde 5 mal wiederholt. Am Ende wurden die Beads in den 1,5 ml Tubes durch Zentrifugieren pelletiert und die Überstände verworfen.

Bei 2.2.1.9 wurden jeweils 200 µl Melt und Annealing Buffer XT pro Tube verwendet; die Volumina bei der Bead-Mengenbestimmung mittels Partikelzähler wurden beibehalten. Die Beadrecovery-Rate errechnet sich aus dem Quotienten der gemessenen und eingesetzten Beads ($1,09 \times 10^6$ Beads pro Tube). Es wurden jeweils 1,5 µl Enrichment-Primer verwendet, sowie 200 µl je Tube Enhancing Fluid zum Waschen und am Ende die Beads in 50 µl Enhancing Fluid resuspendiert. Bei 2.2.1.11 wurden pro 1,5 ml Tube (also pro cpb-Zahl-Titration) 5 µl gewaschene Enrichment-Beads zu den Beads gegeben; das Waschen im MPC erfolgte mit 200 µl Enhancing Fluid. Bei 2.2.1.12 wurden 100 µl Melt verwendet; das Waschen mit Annealing Buffer XT wurde mit 200 µl durchgeführt und die Beads in mindestens 50 µl Annealing Buffer resuspendiert. Es wurden 1,5 µl Sequencing-Primer pro Tube dazugegeben, diese wie beschrieben annealt, die Beads dann in 200 µl Annealing Buffer gewaschen und wie beschrieben ausgezählt.

Die Titrationsergebnisse werden auf Seite 71 ff. dargestellt.

2.2.6.3 Ansatz der emPCR mit den General Librarys

Nach der Titration wurden die vier aqf-Librarys in einer großen emPCR angesetzt (Tab. 6), wobei aqf3b und aqf4b sich einen Cup teilten (Ansatz im gleichen Capture-Bead-Tube) und aqf5b und aqf6b den anderen. Es wurde das unter Abschnitt 2.2.1 beschriebene Protokoll verwendet.

	Library Name	MID	Konzentration in ng/µl	durchschnittl. Länge in bp	Konzentration in Moleküle/µl	Anteil der Beads in %	cpb	Verdünnungen		µl
								in EB	in bidest	Library
Cup A	aqf3b	1	1,971	700	5164858796	25	0,4	1: 100	1: 10	1,36
	aqf4b	2	1,325	700	3472063879	25	0,35	1: 100	1: 10	1,76
Cup B	aqf5b	3	0,456	700	1194914059	25	0,5	1: 100		0,73
	aqf6b	4	0,069	700	180809364	25	0,05	1: 100		0,48

Tab. 6: emPCR-Ansatz der vier aqf General Librarys

„Anteil der Beads in Prozent“ beschreibt den Anteil an Capture-Beads, die von einer Library besetzt werden sollen, bezogen auf die Gesamtmenge der verwendeten Beads des Kits. Das Ergebnis der in der Tabelle durchgeführten Berechnung ist die Menge an Library in µl, die zu den Capture-Beads dazugegeben werden muss. Die durchschnittlichen Längen und Konzentrationen der Librarys wurden den Ergebnissen der DNA- und RNA-Chips entnommen (bzw. bei aqf6b wurde die Konzentration mit dem Qubit gemessen).

Die Ergebnisse des Beadrecovery sind auf Seite 73 gezeigt.

2.2.6.4 Laufstart mit den General Librarys

In Tab. 7 ist die Zusammenstellung der beiden Regionen auf der PTP dargestellt. Es wurde das unter Kapitel 2.2.2 beschriebene Protokoll verwendet.

Region	emPCR-Ansatz	geladene Bead-Menge	Name der Librarys	MID
1	Cup A	1591810	aqf3b	1
			aqf4b	2
	Cup B	608190	aqf5b	3
			aqf6b	4
2	Cup A	2200000	aqf3b	1
			aqf4b	2

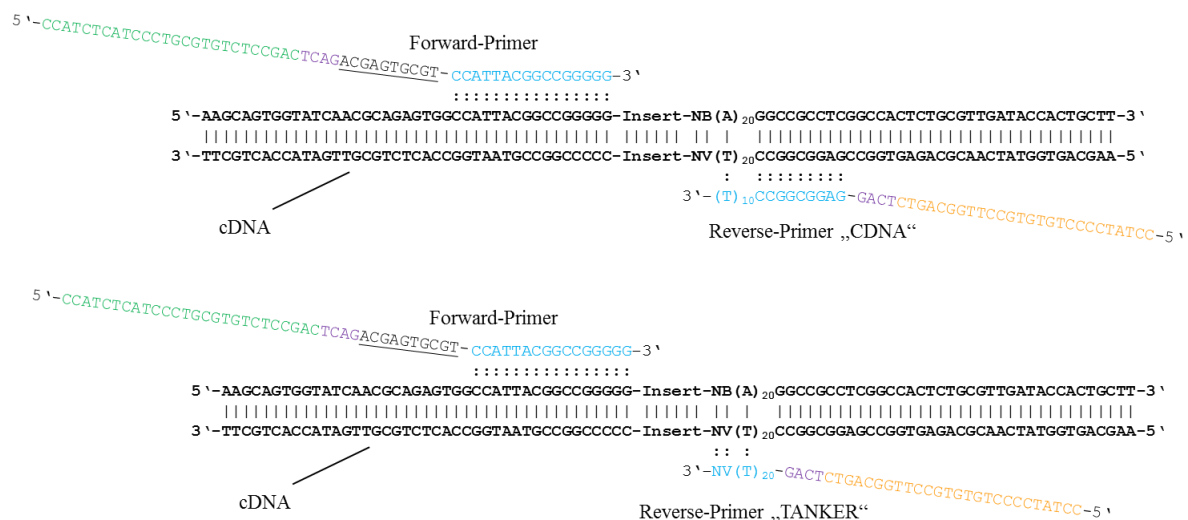
Tab. 7: Zusammensetzung der PTP

Beide Regionen wurden mit jeweils 2,2 Mio. Beads aus der vorangegangenen emPCR beladen. Region 1 bestand also aus zwei unterschiedlichen Bead-Chargen, die jeweils zwei verschiedene Librarys enthielten.

Die Ergebnisse des Runs mit den General Librarys werden auf Seite 74 ff. präsentiert.

2.2.6.5 Erstellung von Amplicon Librarys

Von den aqf-cDNAs aqf3b, aqf4b, aqf5b, aqf6b, aqf9b und aqf10b wurden 454-Librarys nach der Amplicon-Methode hergestellt [83]. In Abb. 36 wird gezeigt, an welchen Bereich der cDNA die für die Herstellung der Amplicon Library verwendeten Primer binden.

**Abb. 36: Bindende Bereiche der Amplicon-Primer**

Die 3'-Enden der Primer, die an die cDNA-Sequenzen binden, sind blau hervorgehoben. Die entsprechenden 454-Adapter-Sequenzen der Primer sind in grün (forward) und orange (reverse), die Key-Sequenzen violett dargestellt; das 10 bp lange MID des Forward-Primers ist unterstrichen (MID1 in diesem Beispiel). Die aqf-cDNAs selber bestehen aus einem Insert, das von den Sequenzen flankiert wird, welche durch das MINT-Kit eingeführt wurden.

Auf der folgenden Seite wird in Tab. 8 gezeigt, welche Primer bei der Herstellung der einzelnen Amplicon Librarys verwendet wurden, Tab. 9 zeigt die Zusammensetzung eines PCR-Ansatzes.

Library	PCR-Ansatz	Forward-Primer	Reverse-Primer
aqf3b	1	ADA-MID1-CDNA	ADB-CDNA
	2		ADB-TANKER
aqf4b	3	ADA-MID2-CDNA	ADB-CDNA
	4		ADB-TANKER
aqf5b	5	ADA-MID3-CDNA	ADB-CDNA
	6		ADB-TANKER
aqf6b	7	ADA-MID4-CDNA	ADB-CDNA
	8		ADB-TANKER
aqf9b	9	ADA-MID5-CDNA	ADB-CDNA
	10		ADB-TANKER
aqf10b	11	ADA-MID6-CDNA	ADB-CDNA
	12		ADB-TANKER

Tab. 8: Primerkombinationen bei der Amplicon Library-Herstellung

Reagenz	Konzentration	Volumen für eine PCR in μl
Forward Primer	100 μM	0,2
Reverse Primer	100 μM	0,2
dNTP's	100 mM (jedes)	0,1
PCR-Puffer	10x	10
MgCl ₂	25 mM	10
Betaine	5 M	10
ddH ₂ O		Rest
Taq	10 Units/ μl	0,97
Pfu	10 Units/ μl	0,025
Template		x
Gesamtvolumen		100

Tab. 9: Zusammensetzung eines PCR-Ansatzes

Die Taq- und Pfu-Polymerasen sind Eigenproduktionen des MPI für molekulare Genetik. Der 10x PCR-Puffer besteht aus 500 mM KCl, 150 mM Tris-HCl pH=9,0 und 0,1% Tween20. ddH₂O ist doppelt destilliertes und autoklaviertes Wasser.

Zur Amplicon Library-Erstellung wurde von jedem PCR-Ansatz (Primer siehe Tab. 8) eine Gradienten-PCR durchgeführt, um eine optimale Annealingtemperatur für die PCR zu ermitteln. Dazu wurde in einer 96er PCR-Platte jeder Ansatz auf 12 PCR-Tubes zu je 100 μl aliquotiert (mit 1 ng cDNA pro PCR-Tube), in 12 weiteren PCR-Tubes wurde für jeden Ansatz eine Negativkontrolle durchgeführt (gleicher PCR-Ansatz ohne Template-DNA) und folgendes PCR-Profil benutzt: 94°C für 5 min, 30x (94°C für 30 s, 50-70°C für 30 s, 68°C für 3 min), 10°C für immer. Anschließend wurden davon je 5 μl auf ein Agarose-Gel aufgetragen. Nach Kontrolle des Gelbildes wurde von jeder Library ein Ansatz zweimal mit SPRI-Beads aufgereinigt (Größencutoff bei 500 bp), in 50 μl EB eluiert und die Konzentration bestimmt.

Die Gelbilder und Ergebnisse der Konzentrationsbestimmung sind auf Seite 78 ff. dargestellt.

2.2.6.6 Titration der Amplicon Librarys

Mit den doppelt SPRI-aufgereinigten Amplifikationsprodukten aus Position 6 der Gradienten-PCR von allen Librarys wurde eine Titration bei je 0,5, 1, 2 und 4 cpb durchgeführt. Der Ablauf der Titration erfolgte wie in Kapitel 2.2.6.2 beschrieben. Tab. 10 zeigt die Berechnung der Titrationsansätze.

Name	MID	Konzentration in ng/μl	durchschnittl. Länge in bp	Konzentration in Moleküle/μl	0,5 cpb	μl an Library für 1 cpb	2 cpb	4 cpb
aqf3b	1	22,75	1000	20865138592,75	2,61E-05	5,22E-05	1,04E-04	2,09E-04
aqf4b	2	14,03	1000	12867599756,32	4,24E-05	8,47E-05	1,69E-04	3,39E-04
aqf5b	3	10,71	1000	9822665245,20	5,55E-05	1,11E-04	2,22E-04	4,44E-04
aqf6b	4	10,28	1000	9428291197,08	5,78E-05	1,16E-04	2,31E-04	4,62E-04
aqf9b	5	19,92	1000	18269607066,71	2,98E-05	5,97E-05	1,19E-04	2,39E-04
aqf10b	6	25,84	1000	23699128845,57	2,30E-05	4,60E-05	9,20E-05	1,84E-04

Tab. 10: Berechnung der Titration der Amplicon Librarys

Aus den nach der doppelten SPRI-Aufreinigung gemessenen Konzentrationen und der anhand der Gelbilder abgeschätzten durchschnittlichen Längen der sechs aqf-Librarys wurde die Konzentration in Molekülen pro Volumen errechnet (vergleiche Kapitel 1.4.5 auf Seite 24). Da bei den doppelsträngigen Amplicon Librarys nur ein Strang an die Capture-Beads aufgrund der Sequenz binden kann, wird nur die Hälfte der Konzentration in ng/μl für die Berechnung der Molekül-Konzentration herangezogen.

Die Titrationsergebnisse der Amplicon Librarys werden auf Seite 81 f. gezeigt.

2.2.6.7 Ansatz der emPCR mit den Amplicon Librarys

Mit den sechs aqf-Amplicon Librarys wurde anhand der Titrations-Ergebnissen eine große emPCR angesetzt. Es wurde das unter 2.2.1 beschriebene Protokoll verwendet, wobei die Mengen und Volumina auf 16,67% heruntergerechnet wurden, damit jede Library in einem separaten Cup angesetzt und in der emPCR amplifiziert werden konnte. Tab. 11 zeigt die Berechnung des emPCR-Ansatzes.

	Library Name	MID	Konzentration in ng/μl	durchschnittl. Länge in bp	Konzentration in Moleküle/μl	Anteil der Beads in %	cpb	Verdünnungen in EB	in bidest	μl Library
Cup A	aqf3b	1	22,75	1000	20865138593	16,67	0,5	1: 100	1: 100	2,80
Cup B	aqf4b	2	14,03	1000	12867599756	16,67	4	1: 100	1: 10	3,63
Cup C	aqf5b	3	10,71	1000	9822665245	16,67	6	1: 100		0,71
Cup D	aqf6b	4	10,28	1000	9428291197	16,67	4	1: 100		0,49
Cup E	aqf9b	5	19,92	1000	18269607067	16,67	0,5	1: 100	1: 100	3,19
Cup F	aqf10b	6	25,84	1000	23699128846	16,67	2	1: 100	1: 10	0,98

Tab. 11: emPCR-Ansatz der aqf-Amplicon Librarys

„Anteil der Beads in Prozent“ beschreibt den Anteil an Capture-Beads, die von einer Library besetzt werden sollen, bezogen auf die Gesamtmenge der verwendeten Beads des Kits. Das Ergebnis der durchgeführten Berechnung ist die Menge an Library in μl, die zu den vorher aliquotierten Capture-Beads dazugegeben werden musste.

Die Ergebnisse des Beadrecovery werden auf Seite 82 f. dargestellt.

2.2.6.8 Laufstart mit den Amplicon Librarys

Auf der folgenden Seite in Tab. 12 ist die Zusammenstellung der beiden Regionen auf der PTP dargestellt. Es wurde das unter Abschnitt 2.2.2 beschriebene Protokoll verwendet.

Region	emPCR-Ansatz	geladene Bead-Menge	Name der Librarys	MID
1	Cup A	531200	aqf3b	1
	Cup B	700000	aqf4b	2
	Cup C	700000	aqf5b	3
2	Cup D	700000	aqf6b	4
	Cup E	700000	aqf9b	5
	Cup F	700000	aqf10b	6

Tab. 12: Beladung beider PTP-Regionen des Amplicon-Laufs

Es wurden von jeder Library 0,7 Mio. Beads aus der vorangegangenen emPCR geladen (mit Ausnahme von aqf3b, da hier nicht genügend Beads in der emPCR produziert worden waren), wobei jede Region mit Beads von je drei Librarys beladen wurde.

Die Runergebnisse sind auf Seite 83 ff. gezeigt.

2.2.6.9 Amplifikations-Test mit den aqf-Amplicon Librarys

Nach dem fehlgeschlagenen Run mit den Amplicon Librarys wurde mit diesen ein Amplifikationstest durchgeführt. Dabei wurden Primer verwendet, die die Adapter-A- und -B-Sequenz tragen. Tab. 13 zeigt die Zusammensetzung der jeweiligen PCR-Ansätze. Neben den sechs Amplicon Librarys wurde eine Negativkontrolle ohne Template-DNA angesetzt und die PCR mit folgendem Programm gefahren: 4 min bei 94°C; 50x (30 s bei 94°C; 4,5 min bei 58°C; 30 s bei 68°C); anschließend bei 10°C halten.

Reagenz	Konzentration	Volumen für eine PCR in µl
454PrimerA	100 µM	0,2
454PrimerB	100 µM	0,2
dNTP's	100 mM (jedes)	0,1
PCR-Puffer	10x	10
MgCl ₂	25 mM	10
Betaine	5 M	10
ddH ₂ O		Rest
Taq	10 Units/µl	0,97
Pfu	10 Units/µl	0,025
Template		0,5
Gesamtvolumen		100

Tab. 13: PCR-Ansatz des Amplifikations-Tests

Als Template diente die Amplicon Library-DNA; Taq- und Pfu-Polymerasen sind Eigenproduktionen des MPI für molekulare Genetik. Der 10x PCR-Puffer besteht aus 500 mM KCl, 150 mM Tris-HCl pH=9,0 und 0,1% Tween20. ddH₂O ist doppelt destilliertes und autoklaviertes Wasser.

Das Gelbild des Amplifikationstests ist auf Seite 86 dargestellt.

2.2.6.10 Amplifikations-Tests mit aqf-Sanger-Klonen

Zur Fehlersuche wurden mit aqf-cDNA-Sanger-Klonen 454-Amplicon Librarys hergestellt (sie werden Sanger-Klone genannt, weil sie ursprünglich für die Sanger-Sequenzierung erzeugt wurden). Die cDNA-Klone wurden vom Labor Kube zur Verfügung gestellt (die aqf-cDNA wurde in den Vektor pUC19 ligiert, mittels Elektroporation in elektrokompente *E.coli*-Zellen transformiert, vereinzelt, die einzelnen Klone gepickt und die DNA isoliert, welche daher also die gleiche Insert-DNA pro Klon trägt) und besaßen eine Insertgröße um die 1 kb. Als Test wurden aus fünf Sanger-Klonen in einer Gradienten-PCR Amplicon Librarys erstellt, wobei als Primer ADA-MID1-CDNA und ADB-CDNA

verwendet wurden; die Zusammensetzung des PCR-Mixes ist aus Tab. 9 auf Seite 50 ersichtlich, es wurden pro 100 µl PCR-Ansatz ein Template-Menge von 20 ng verwendet und mit folgendem Programm amplifiziert: 94°C für 5 min, 30x (94°C für 30 s, 50-70°C für 30 s, 68°C für 3 min), 10°C für immer. Gleichzeitig wurde noch eine Negativ-Kontrolle ohne Template-DNA mit in der PCR gefahren. Anschließend wurden von jedem 100 µl-PCR-Ansatz je 5 µl auf ein Agarose-Gel zur Kontrolle aufgetragen. Von jedem aqf-cDNA-Sanger-Klon-Ansatz der Gradienten-PCR wurden jeweils 0,5 µl als Template-DNA in einer neuen PCR angesetzt, deren Zusammensetzung in Tab. 13 ersichtlich ist. Zusätzlich wurden noch eine Negativ-Kontrolle (PCR-Ansatz ohne Template-DNA) und eine Positiv-Kontrolle mit folgendem Programm gefahren: 4 min bei 94°C; 30x (30 s bei 94°C; 30 s bei 58°C; 3 min bei 68°C); anschließend bei 10°C halten. Bei der Positiv-Kontrolle wurden 20 ng je 100 µl-PCR-Ansatz einer genomischen 454-Shotgun-Library verwendet, die im 454 gute Sequencing-Ergebnisse lieferte.

Alternativ dazu wurde mittels Two-Step-PCR versucht, aus einem Sanger-Klon eine Amplicon Library zu erstellen. In einer ersten Gradienten-PCR wurden die Shortfor- und Shortrev-Primer verwendet (Zusammensetzung des PCR-Mixes siehe Tab. 9 auf Seite 50) mit 20 ng Template pro Ansatz und mit folgendem PCR-Programm gefahren: 94°C für 5 min, 30x (94°C für 30 s, 50-70°C für 30 s, 68°C für 3 min), 10°C für immer. Eine entsprechende Negativ-Kontrolle ohne Template wurde angesetzt. Nach der PCR wurden 5 µl pro Ansatz auf ein Agarose-Gel aufgetragen. In einer zweiten PCR (Zusammensetzung des PCR-Mixes siehe Tab. 9 auf Seite 50) wurden je 1 µl der ersten PCR als Template-DNA verwendet, und mit den ADA-MID1-CDNA- und ADB-CDNA-Primern mit dem gleichen PCR-Programm amplifiziert (eine Negativkontrolle ohne Template wurde ebenfalls angesetzt). Als Positivkontrolle diente eine genomische 454-Library (wobei andere Primer verwendet wurden: 454PrimerA und 454PrimerB). Von den Ansätzen wurden 5 µl auf ein Agarosegel aufgetragen.

Die Gelbilder der Tests mit den Sanger-Klonen werden auf Seite 87 ff. gezeigt.

2.2.7 Entwicklung einer neuen Methode zur gerichteten Sequenzierung

Da die vorangegangenen Versuche, die cDNAs über die Erstellung von General- und Amplicon Librarys zu sequenzieren, scheiterten und keine brauchbaren Reads in entsprechenden Ausbeuten brachten, wurde eine neue Methode zur Erstellung von 454-Librarys erdacht, die eine gerichtete Sequenzierung ermöglicht. Das Prinzip der Library-Herstellung ist in Abb. 37 dargestellt, anhand dessen eine erste Test-Library erprobt und später bei Erfolg alle entsprechenden cDNAs sequenziert werden sollen.

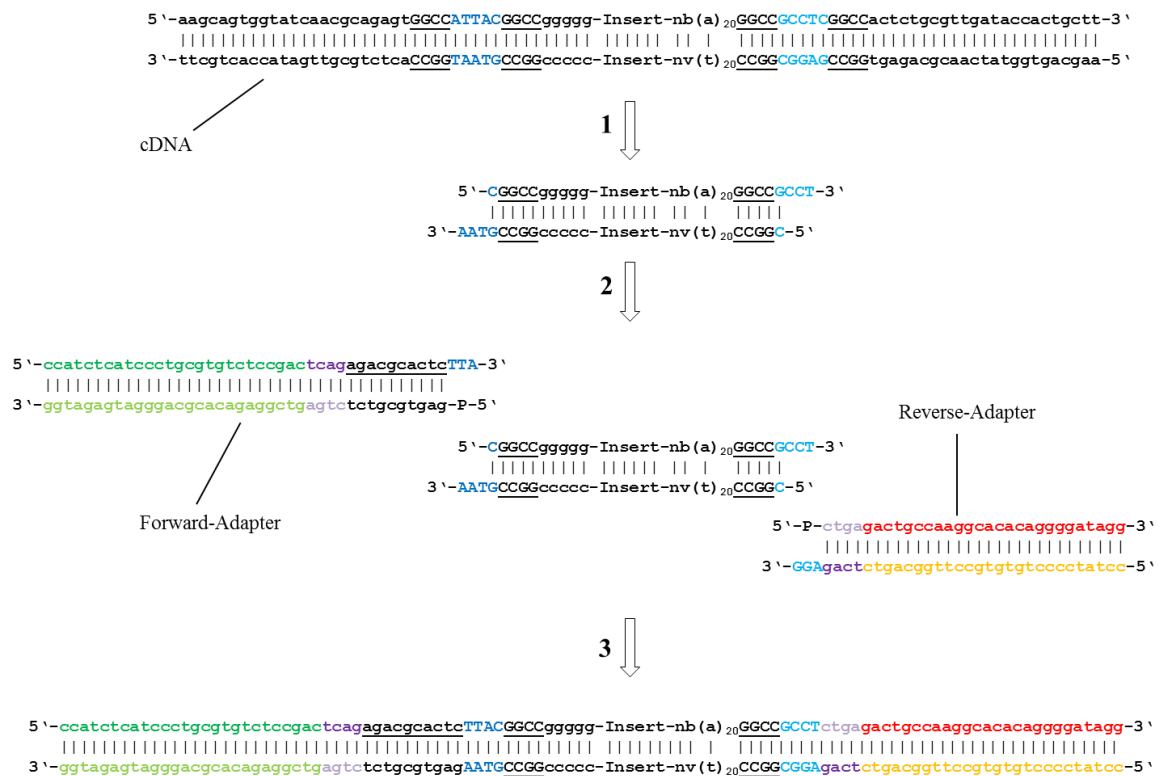


Abb. 37: Erstellung von 454-Librarys mittel gerichteter Ligation

Die mittels MINT-Kit hergestellte cDNA besitzt Schnittstellen für das Restriktionsenzym SfiI (in Großbuchstaben dargestellt), die SfiI-Erkennungssequenzen sind unterstrichen. Nach einem Restriktionsverdau (1) und Entfernen der 5'- und 3'-Sequenzen besitzt die cDNA sogenannte „Sticky“-Ends, wobei diese von unterschiedlicher Sequenz sind (dunkel- und hellblau), welche eine gerichtete Ligation und damit gerichtete Sequenzierung zulassen. Es werden Adapter konstruiert, die einerseits die 454-Adapter-A- und -B-Sequenzen (grün und rot), Key (violett) und MID (unterstrichen) tragen, und andererseits ebenfalls zu den Enden der geschnittenen cDNA passende Sticky-Ends aufweisen. Zudem müssen die 5'-Enden der Adapter, welche an die cDNA ligiert werden sollen, phosphoryliert sein („P“). Nach Ligation und Aufreinigung zum Entfernen unverbrauchter Adapter (mittels SPRI) sollte man sequenzierbare Librarys aus der cDNA erhalten (3).

2.2.7.1 Erstellung einer Test-Library

Von der aqf65-cDNA wurden 5 µg in einem Restriktionsansatz geschnitten: zu 5 µg cDNA gelöst in 19,06 µl EB-Puffer wurden in ein 0,2 ml PCR-Tube 1 µl 100x BSA, 10 µl 10x NEB-Puffer 2, 3 µl SfiI (60 U) gegeben und mit 66,94 µl doppelt destilliertem und autoklaviertem Wasser auf ein Gesamtvolumen von 100 µl aufgefüllt. Nach Vortexen erfolgte die Inkubation bei 50°C für eine Stunde in einem Thermocycler. Anschließend erfolgte eine Aufreinigung des Restriktionsansatzes mit dem

MinElute PCR Purification Kit nach den Angaben des Herstellers, wobei 500 µl PBI-Puffer verwendet wurden und mit 11 µl EB-Puffer eluiert wurde.

Aus den Oligonukleotiden „AdapterB-SfiB“ und „AdapterB-primelSfi“ wurde der Reverse-Adapter, aus „AdapterA-MID3-SfiA“ und „AdapterA-MID3-primelSfi“ wurde der Forward-Adapter wie in Abschnitt 2.2.5 auf Seite 44 beschrieben hergestellt. Für die Ligation wurden zu den 11 µl der aufgereinigten Restriktion in einem 0,2 ml PCR-Tube 1 µl 5 mM ATP-Lösung, 5 µl der hergestellten Adapter, 2 µl 10x Ligase-Puffer und 1 µl verdünnte T4-Ligase gegeben -verdünnte Ligase bestand aus 1µl Ligase (1 U) und 3 µl 1x-Ligase Puffer, gevortext und in einem Thermocycler mit folgendem Programm inkubiert: 16 Stunden bei 4°C, 10 Minuten bei 65°C, Halten bei 10°C. Danach wurden die 20 µl Ligationsansatz mit dem MinElute PCR Purification Kit nach den Angaben des Herstellers aufgereinigt, wobei 100 µl PBI-Puffer verwendet wurden und mit 100 µl EB-Puffer eluiert wurde. Mit den 100 µl des aufgereinigten Ligationsansatzes erfolgte eine doppelte Aufreinigung mit SPRI-Beads wie unter Abschnitt 2.2.3 auf Seite 43 beschrieben mit einem Cutoff bei 500 bp; am Ende erfolgte die Elution in 20 µl EB Puffer. Die DNA-Konzentration wurde mit einem Fluorometer (Qubit) bestimmt.

Zur Kontrolle wurde mit 0,5 µl der erhaltenen Library eine Test-PCR mit dem in Tab. 13 auf Seite 52 beschriebenen Ansatz durchgeführt, wobei eine entsprechende Negativkontrolle ohne Template-DNA mitgefahren wurde. Jeweils 5 µl davon wurden auf ein Agarose-Gel aufgetragen (Seite 91).

2.2.7.2 emPCR mit der Test-Library

Die Test-Library wurde mit anderen Librarys in einer emPCR angesetzt (Tab. 14), wobei das unter 2.2.1 beschriebene Protokoll verwendet wurde. Die anderen Librarys wurden alle nach dem General-Library-Protokoll hergestellt

Library Name	MID	Konzentration in ng/µl	durchschnittl. Länge in bp	Konzentration in Moleküle/µl	Anteil der Beads in %	cpb	Verdünnungen			µl Library
							in EB	in bidest		
271-15	4	6,197	1000	11367144076	47,1	0,5	1: 100	1: 10		1,45
mge011	1	0,702	1000	1287677125	0,4	0,7	1: 100	1: 100		1,52
Test-Library	3	62,3	1000	57138379531	1,5	1	1: 100	1: 100	1: 10	1,84
FR7	7	19,7	500	72271337192	0,5	0,25	1: 100	1: 100	1: 100	1,21
FR8	8	16,075	500	58972677429	0,5	0,25	1: 100	1: 100	1: 100	1,48

Tab. 14: emPCR-Ansatz mit der Test-Library

„Anteil der Beads in Prozent“ beschreibt den Anteil an Capture-Beads, die von einer Library besetzt werden sollen; die Summe ist hier 50%, also die Hälfte, da nur ein Cup der emPCR dargestellt ist. D.h. die Library 271-15 besitzt einen Anteil von 94,2% An den Capture-Beads des Cups, in dem die in der Tabelle aufgeführten Librarys zusammen pipettiert wurden. Die Librarys 271-15 und mge011 sind in vorangegangenen Läufen schon sequenziert worden, weshalb der cpb-Wert bei diesen schon eingestellt ist. Bei FR7, FR8 und der Test-Library wurde der cpb-Wert einfach geschätzt.

Die Ergebnisse des Beadrecovery sind auf Seite 92 dargestellt.

2.2.7.3 Laufstart mit der Test-Library

Es wurde das unter Abschnitt 2.2.2 beschriebene Protokoll verwendet, wobei 2,2 Mio. an aufgereinigten Beads (vergleiche Tab. 28 auf Seite 92) des emPCR-Ansatzes (Tab. 14) auf Region 1 der PTP geladen wurden. Die Runergebnisse werden auf Seite 92 ff. gezeigt.

2.2.7.4 Erstellen von Librarys mit der neuen Methodik aus aqf-cDNA

Aus den aqf cDNAs wurden dem unter Abschnitt 2.2.7.1 beschriebenen Protokoll folgend 454-Librarys nach der neuen Methodik hergestellt, einschließlich der dort beschriebenen Test-PCR. In Tab. 15 sind die verwendeten Oligonukleotide zur Herstellung der Adapter und die zugeordneten MIDs aufgeführt.

cDNA	Oligos für Adapter A	MID	cDNA	Oligos für Adapter A	MID
aqf3b	AdapterA-MID20-SfiIA AdapterA-MID20-primelSfi	20	aqf15	AdapterA-MID20-SfiIA AdapterA-MID20-primelSfi	20
aqf4b	AdapterA-MID21-SfiIA AdapterA-MID21-primelSfi	21	aqf25	AdapterA-MID21-SfiIA AdapterA-MID21-primelSfi	21
aqf5b	AdapterA-MID22-SfiIA AdapterA-MID22-primelSfi	22	aqf35	AdapterA-MID22-SfiIA AdapterA-MID22-primelSfi	22
aqf6b	AdapterA-MID23-SfiIA AdapterA-MID23-primelSfi	23	aqf45	AdapterA-MID23-SfiIA AdapterA-MID23-primelSfi	23
aqf9b	AdapterA-MID24-SfiIA AdapterA-MID24-primelSfi	24	aqf55	AdapterA-MID24-SfiIA AdapterA-MID24-primelSfi	24
aqf10b	AdapterA-MID25-SfiIA AdapterA-MID25-primelSfi	25	aqf65	AdapterA-MID25-SfiIA AdapterA-MID25-primelSfi	25

Tab. 15: Verwendete Adaptersequenzen

Es sind für die aqf-cDNAs nur die verwendeten Oligonukleotide für Adapter A angegeben, da die Adapter B-Sequenzen bzw. Oligonukleotide die gleichen sind wie in Kapitel 2.2.7.1 beschrieben.

Das Gelbild der Test-PCR wird auf Seite 95 f. dargestellt.

2.2.7.5 emPCR der ersten sechs aqf-Librarys

Die Librarys aqf3b, aqf4b, aqf5b, aqf6b, aqf9b und aqf10b wurden in einer halben emPCR angesetzt (Tab. 16) für eine Region. Die andere Hälfte der emPCR für die zweite Region des Sequencing-Runs bestehend aus General Librarys wird nicht dargestellt. Es wurde das unter Kapitel 2.2.1 beschriebene Protokoll verwendet.

Library Name	MID	Konzentration in ng/μl	durchschnittl. Länge in bp	Konzentration in Moleküle/μl	Anteil der Beads in %	cpb	Verdünnungen in EB	Verdünnungen in bidest	μl Library
aqf3b	20	37,4	1000	34301370698	8,33333333	1,5	1: 100	1: 100	2,55
aqf4b	21	38,4	1000	35218519647	8,33333333	1,5	1: 100	1: 100	2,48
aqf5b	22	35,2	1000	32283643009	8,33333333	1,5	1: 100	1: 100	2,71
aqf6b	23	8,4	1000	7704051173	8,33333333	1,5	1: 100	1: 10	1,14
aqf9b	24	28,5	1000	26138745050	8,33333333	1,5	1: 100	1: 100	3,35
aqf10b	25	29,7	1000	27239323789	8,33333333	1,5	1: 100	1: 100	3,21

Tab. 16: emPCR-Ansatz von aqf3b bis aqf10b

„Anteil der Beads in Prozent“ beschreibt den Anteil an Capture-Beads, die von einer Library besetzt werden sollen; die Summe ist auch hier wieder 50%, also die Hälfte, da nur ein Cup der emPCR dargestellt ist. Die cpb-Werte sind anhand des Ergebnisses der Test-Library von 1 auf 1,5 erhöht worden.

Die Ergebnisse des Beadrecovery sind auf Seite 96 gezeigt.

2.2.7.6 Laufstart mit den ersten sechs aqf-Librarys

Es kam das unter Abschnitt 2.2.2 beschriebene Protokoll zur Anwendung, wobei 2,2 Mio. an aufgereinigten Beads (Tab. 33 auf Seite 96) vom emPCR-Ansatz (Tab. 16) auf Region 1 der PTP geladen wurden. Die Runergebnisse sind auf Seite 97 ff. dargestellt.

2.2.7.7 emPCR mit weiteren sechs aqf-Librarys

Die Librarys aqf15, aqf25, aqf35, aqf45, aqf55 und aqf65 wurden wie bei der vorherigen emPCR in einer halben emPCR angesetzt (Tab. 17). Die andere Hälfte der emPCR bestand aus General Librarys und wird nicht dargestellt. Es wurde das in Kapitel 2.2.1 beschriebene Protokoll verwendet.

Library Name	MID	Konzentration in ng/μl	durchschnittl. Länge in bp	Konzentration in Moleküle/μl	Anteil der Beads in %	cpb	Verdünnungen		μl Library
							in EB	in bidest	
aqf15	20	35,6	1000	32650502589	8,33333333	1,3	1: 100	1: 100	2,32
aqf25	21	30,4	1000	27881328054	8,33333333	1,3	1: 100	1: 100	2,72
aqf35	22	37,2	1000	34117940908	8,33333333	1,3	1: 100	1: 100	2,22
aqf45	23	30,6	1000	28064757843	8,33333333	1,3	1: 100	1: 100	2,70
aqf55	24	26,9	1000	24671306732	8,33333333	1,3	1: 100	1: 100	3,07
aqf65	25	28,3	1000	25955315260	8,33333333	1,3	1: 100	1: 100	2,92

Tab. 17. emPCR-Ansatz von aqf15 bis aqf65

„Anteil der Beads in Prozent“ beschreibt den Anteil an Capture-Beads, die von einer Library besetzt werden sollen; da nur ein Cup der emPCR dargestellt ist, ist die ihre Summe 50%. Die cpb-Werte sind anhand des Ergebnisses des Sequencing-Runs der ersten sechs aqf-Librarys auf 1,3 verringert worden.

Die Beadrecovery-Ergebnisse werden auf Seite 102 gezeigt.

2.2.7.8 Laufstart mit den weiteren aqf-Librarys

Es kam das unter Abschnitt 2.2.2 beschriebene Protokoll zur Anwendung, wobei 2,2 Mio. an aufgereinigten Beads (Tab. 37 auf Seite 102) vom emPCR-Ansatz (Tab. 17) auf Region 1 der PTP geladen wurden. Die Ergebnisse des Sequencing-Runs werden auf Seite 102 ff. dargestellt.

2.2.7.9 Erweiterung auf andere Templates (Sequenzierung einer Illumina-Library)

Um auch aus Template-DNA mit der neuen Methodik Librarys erstellen zu können, die keine SfiI-Schnittstellen tragen, sollen diese in einer PCR an die Template-DNA heranamplifiziert werden. Die Template-DNA, „OCon“ genannt, besaß eine durchschnittliche Länge von 250 bp (Angaben des Auftraggebers) und eine Konzentration von 290 ng/μl. Sie war ursprünglich eine fertige Illumina-Library, die aufgrund von zeitlichen Engpässen auf dem 454 sequenziert werden sollte. Da die Sequenzrichtung der OCon-Illumina-Library auf dem 454 beibehalten werden sollte, kam die neue Methodik zur Anwendung. Das 5'-Ende des Forward-Stranges der OCon-DNA besaß die Sequenz 5'-CTCTGGACCTTGGCTGTCACTCAGTT-3', das 5'-Ende des Reverse-Stranges die folgende: 5'-CCTTGGCTGTCACTCACTGCG-3'. Das Prinzip der 454-Library-Herstellung ist in Abb. 38 dargestellt, Tab. 18 zeigt die Zusammensetzung des PCR-Ansatzes (folgende Seite).

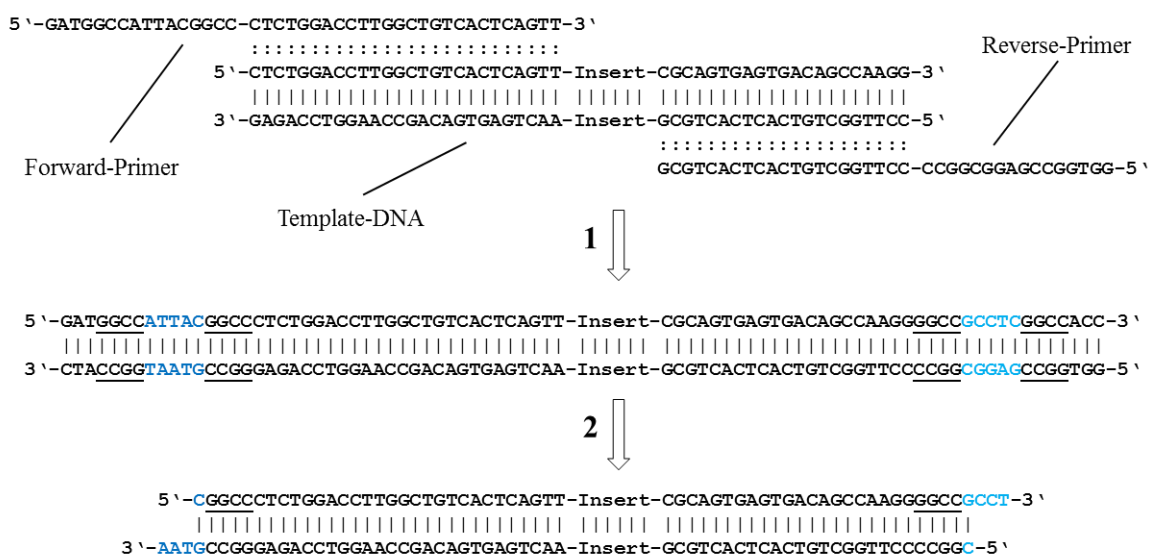


Abb. 38: Heranamplifizieren der SfiI-Sites

Mit der Template-DNA wird eine PCR durchgeführt, in der die angegebenen Forward- und Revers-Primer verwendet werden, welche an ihren 5'-Enden 16 bp-lange Overhangs besitzen (1). Durch diese tragen die Amplifikationsprodukte nach der PCR SfiI-Erkennungssequenzen (unterstrichene Bereiche). Diese stellen den Ausgangspunkt für die Library-Herstellung mit der neuen Methodik dar, da sie von SfiI geschnitten werden (2) und dann die für die gerichtete Ligation benötigten Sticky Ends ergeben (dunkel- und hellblau). Vergleiche mit Abb. 37 auf Seite 54.

Reagenz	Konzentration	Volumen für eine PCR in µl
SfiIA-OConFor	100 µM	0,1
SfiIB-OConRev	100 µM	0,1
dNTP's	100 mM (jedes)	0,1
PCR-Puffer	10x	10
MgCl ₂	25 mM	10
Betaine	5 M	10
ddH ₂ O		63,7
Taq	10 Units/µl	0,97
Pfu	10 Units/µl	0,025
Template	1450 ng	5
Gesamtvolumen		100

Tab. 18: PCR-Ansatz zum Heranamplifizieren der SfiI-Sites

Taq- und Pfu-Polymerasen sind Eigenproduktionen des MPI für molekulare Genetik. Der 10x PCR-Puffer besteht aus 500 mM KCl, 150 mM Tris-HCl pH=9,0 und 0,1% Tween20. ddH₂O ist doppelt destilliertes und autoklaviertes Wasser.

Folgendes Programm wurde bei der PCR verwendet: 4 min bei 94°C, dann 5x (94°C für 30 s, 58°C für 1 min, 68°C für 1 min), dann Halten bei 10°C. Danach erfolgte eine Aufreinigung der PCR mit dem MinElute PCR Purification Kit nach den Angaben des Herstellers, wobei 500 µl PBI-Puffer verwendet wurden und 86 µl EB-Puffer zur Elution. Mit den 86 µl Eluat wurde eine 454-Library wie unter Abschnitt 2.2.7.1 beschrieben hergestellt, unter Verwendung der Oligonukleotide AdapterB-SfiIB, AdapterB-primelSfi, AdapterA-MID26-SfiIA und AdapterA-MID26-primelSfi. Folgende Änderungen

bzw. Anpassungen fanden statt: die beiden SPRI-Aufreinigung wurden mit einer Menge Beads durchgeführt, die einen Cutoff der Charge bei 200 bp ermöglicht.

Die so erhaltene Library wurde einer emPCR untergemischt und mit den aus dem Beadrecovery erhaltenen Beads ein Sequenzierlauf gestartet, wobei die unter Kapitel 2.2.1 und 2.2.2 beschriebenen Protokolle zur Anwendung kamen.

Die Ergebnisse des Sequencing-Runs mit der OCon-Library sind auf Seite 107 ff. gezeigt.

2.2.7.10 Antikörpersequenzierung von schweren Immunglobulinketten

Die Arbeitsgruppe Konthur am MPI für molekulare Genetik wollte die Diversitäten der Antikörperrepertoires verschiedener Spender unterschiedlichen Alters und Gesundheitszustandes bestimmen. Die hohe Diversität des humanen Antikörperrepertoires ergibt sich aufgrund des V(D)J-Rekombinationsmechanismus unterschiedlicher V-, D- und J-Gene der leichten und schweren Immunglobulinketten [116]. Für die Sequenzierung der schweren Immunglobulinketten (IgH) einschließlich eines kurzen Abschnittes des konstanten Bereiches werden Leselängen von ungefähr 400 bp benötigt. Daher kommt das 454-Pyrosequencing als einzige Sequenziermethode derzeit infrage; andere NGS-Systeme produzieren zu kurze Leselängen und die Sangersequenzierung ist zu teuer.

Bis jetzt basieren alle publizierten Antikörpersequenzierungen auf sequenzspezifischen Amplifikationen per PCR des Antikörperrepertoires: ein Primer bindet innerhalb des V-Gen-Bereichs, der andere entweder innerhalb des J-Gen- oder des konstanten Bereichs. Diese Herangehensweise hat den Nachteil, dass es zu einem Bias gegenüber bestimmter Sequenzen und gleichzeitig zu einer Limitierung der Amplifikation bestimmter V-Gene kommen kann [117-120]. Um diese Probleme zu umgehen, wurde von Florian Rubelt (Arbeitsgruppe Konthur, MPI für molekulare Genetik, Berlin) eine alternative Vorgehensweise bei der cDNA-Synthese gewählt. Dabei verwendete er einen PlugOligo-Adapter (aus dem MINT-Kit) für das 5'-Ende in Verbindung mit 3'-Primersequenzen, die auf die konservierte CH1-Domäne aufsetzen. Somit war er bei der cDNA-Herstellung unabhängig von Sequenzen der variablen V- oder J-Gene. Die CH1-spezifischen Primer wurden von Florian Rubelt so gewählt, dass die resultierenden Sequenzen in ihre fünf Isotypen bzw. 9 Subtypen aufgeteilt werden können (IgA1, -A2, -D, -E, -G1, -G2, -G3, -G4, -M), um ein Deepsequencing der Antikörperrepertoires verschiedener Spender zu ermöglichen.

Bei der RNA handelte es sich um gesamt-RNA aus peripheren Blutzellen, die während der Erststrangsynthese mittels Oligo-dT-Primer in sscDNA überschrieben wird. Da aber nicht die gesamte cDNA sequenziert werden sollte, sondern nur die IgH-cDNA, fiel eine 454-Sequenzierung der total-cDNA über Rapid- oder General-Library-Protokolle aus. Zudem sollte von den CH1-Bereichen ausgerichtet sequenziert werden. Die 454-Adaptersequenzen mussten also an die cDNA heranamplifiziert werden. Da Versuche zur Erstellung einer Amplicon Library nach Roche-Standard-Protokoll fehl schlugen (Durchführung Florian Rubelt, Ergebnisse nicht gezeigt), sollte zur Antikörpersequenzierung auf die neue Methodik zurückgegriffen werden. Dafür sollten die für die Library-Herstellung benötigten

SfiI-sites nach dem in 2.2.7.9 auf Seite 57 f beschriebenen Prinzip in das Template eingeführt werden. Der schematische Ablauf ist in Abb. 39 auf Seite 61 dargestellt.

Im Folgenden wird die Durchführung für die Vorbereitung eines Runs gezeigt, bei dem eine Region mit vier Antikörper-Librarys sequenziert wurde.

Libraryherstellung. Nach Erststrangsynthese, Heranamplifizieren der SfiI-Sites in Emulsions-PCR [121] und einem Gelausschnitt der Amplifikate (Durchführung dieser drei Schritte von Florian Rubelt, AG Konthur, MPI für molekulare Genetik), wurden die 454-Librarys wie beschrieben hergestellt (Kapitel 2.2.7.1), wobei allerdings die Eingangsmenge an cDNA auf mindestens 200 ng herabgesenkt wurde und Adaptersequenzen mit den Basen „GACT“ als Key verwendet wurden, da es bei Roche inzwischen eine Umstellung auf diesen Key im Zuge der Rapid-Library-Einführung gab. Zur Qualitätskontrolle wurden die Librarys auf einem Agilent Bioanalyzer 2100 High Sensitivity DNA Chip aufgetragen. Die Ergebnisse des Chips sind auf Seite 111 f dargestellt.

emPCR, Beadrecovery und Sequencing. Zur cpb-Zahl-Bestimmung wurden die Librarys in Runs mit anderen Proben untergemischt und eingestellt. Danach wurden vier Librarys in einer emPCR angesetzt (zu gleichen Anteilen) und nach erfolgtem Beadrecovery ein Sequenzierlauf gestartet (auf einer PTP-Region), wobei die unter Kapitel 2.2.1 und 2.2.2 beschriebenen Protokolle zur Anwendung kamen. Die Ergebnisse der Region des Sequencing-Runs sind auf Seite 111 ff gezeigt.

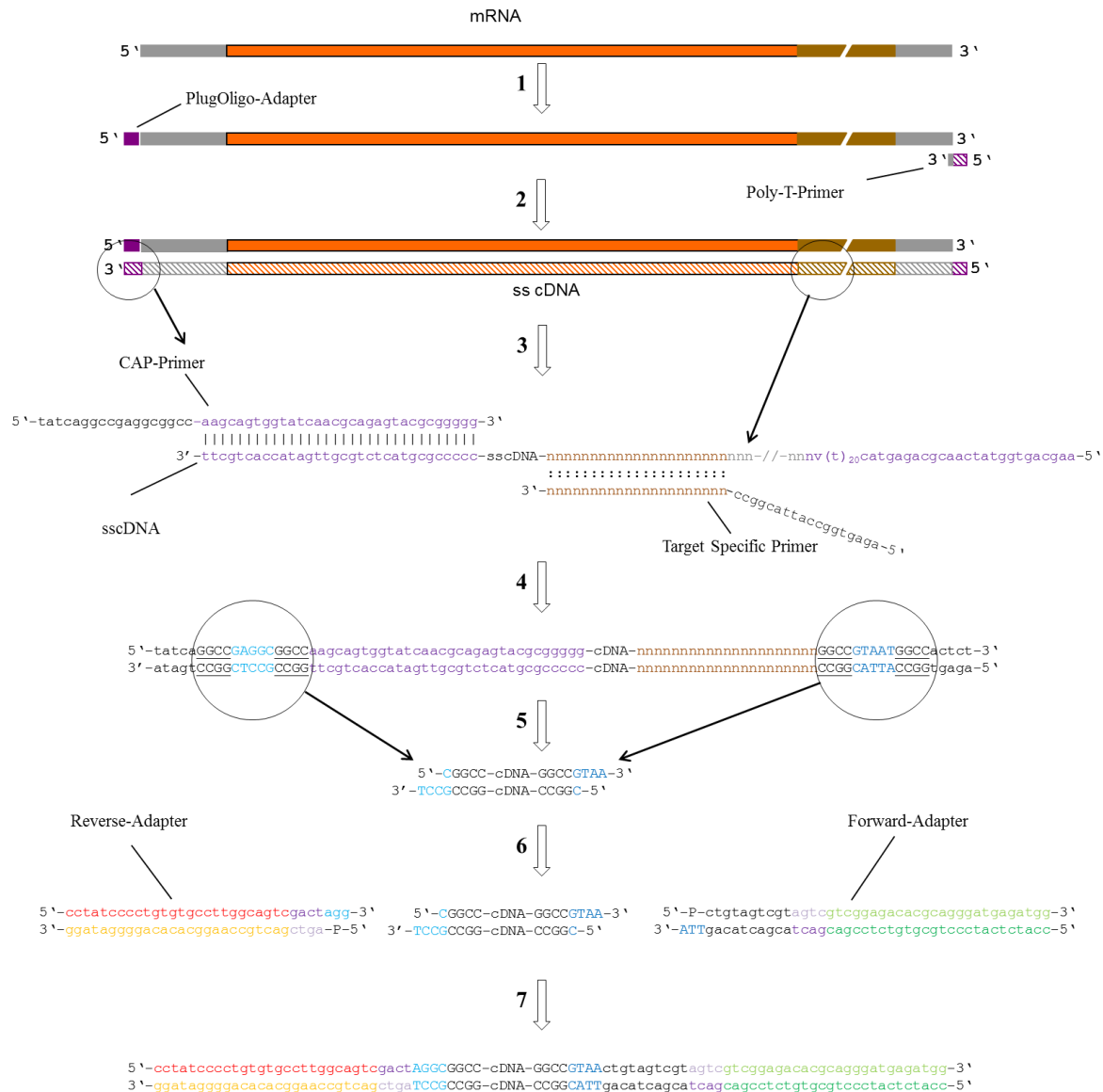


Abb. 39: Prinzip der Antikörpersequenzierung

Dargestellt ist eine IgH-mRNA bestehend aus einer 150 bp langen 5'-untranslated Region (grau), gefolgt von der ca. 380 bp VDJ-Region (orange), der ca. 1500 bp C-Region (braun) und einer 3'-untranslated Region (grau) am Ende. Für die Sequenzierung von IgH-cDNAs wird die gesamte-RNA zur Erststrangsynthese mit einem Poly-T-Primer und einem PlugOligo-Adapter (beide violett) inkubiert (1). Ein Poly-T-Primer bindet den Poly-A-Schwanz der mRNAs, und die Reverse Transkriptase beginnt mit der Synthese des cDNA-Stranges. Kommt sie am 5'-Ende der RNA an, so beginnt sie mit der Addition von einigen Basen –hauptsächlich von Cytosin- an das 3'-Ende der cDNA, die nicht Bestandteil der Template-RNA sind. Auf diesen C-Stretch setzt der PlugOligo-Adapter des MINT-Kits auf, um den die einzelsträngige cDNA (sscDNA) während der Erststrangsynthese verlängert wird (2). Zur spezifischen Amplifikation von IgH-Sequenzen aus den sscDNAs heraus kommen der CAP-Primer und ein Target-Specific-Primer, der innerhalb der CH1-Domäne bindet, zum Einsatz (3). Über die Primer werden auch die SfiI-Sites in das Template eingeführt, dem Prinzip aus Kapitel 2.2.7.9 folgend. Die Amplifikation findet in einer Emulsions-PCR statt [121], um einen möglichen Bias zu verringern. Das Ergebnis der Emulsions-PCR (4) ist ein IgH-cDNA-Template mit SfiI-Restriktion-Sites (Großbuchstaben, Erkennungssequenzen sind unterstrichen) mit einer Länge je nach Template zwischen 600 und 850 bp. Es folgen Restriktion (5) und Ligation (6,7) der 454-Adapter (vergleiche Abb. 37 auf Seite 54). Anhand der Orientierung der so hergestellten 454-Library wird erkennbar, dass die Library ausgehend von der CH1-Domäne im C-Bereich in Richtung der 5'-untranslated Region sequenziert wird (Adapter-A-Sequenzen sind dunkelgrün).

2.2.7.11 Verbesserungen des Protokolls für die Antikörpersequenzierung

Für die Herstellung von 454-Librarys nach der neuen Methodik wurden bisher Insert-DNA-Mengen von 5 µg bis hinunter zu 200 ng verwendet. Da manche Templates aber nicht in derartigen Mengen hergestellt werden können, musste das Protokoll verändert und angepasst werden. Dies galt insbesondere für die Immunglobulin-Librarys, die in Zusammenarbeit mit der AG Konthur erstellt wurden (siehe vorhergehenden Abschnitt); das Protokoll zur 454-Library-Generation sollte an wesentlich kleinere Mengen angepasst werden.

Libraryherstellung. Die Erstellung der cDNAs aus der Gesamt-RNA erfolgte nach den gleichen Prinzipien des vorhergehenden Kapitels (Abb. 39), nur dass der Gelausschnitt entfiel. Die in der Restriktion eingesetzten DNA-Mengen wurden mit dem Qubit bestimmt. Für die Restriktion wurden zu 43,5 µl cDNA 5 µl NEB-Buffer 4, 0,5 µl 100x BSA und 1 µl SfiI (20 U) gegeben, gevortext und für eine Stunde bei 50°C in einem Thermocycler inkubiert. Zur Inaktivierung des Restriktionsenzym SfiI wurde anschließend zu dem Ansatz 1 µl Proteinase K (0,6 mAU) gegeben und mit folgendem Temperaturprofil im Thermocycler inkubiert: 15 min bei 50°C, dann 15 min bei 75°C, Halten bei 10°C. Für die Ligation wurden zu dem Ansatz 1 µl 5 mM ATP-Lösung, 0,1 µl selbst hergestellter Adapter, 2,93 µl 10x Ligase-Puffer, 2,93 µl doppelt destilliertes Wasser und 1 µl verdünnte T4-Ligase (bestehend aus 1µl Ligase (1 U) und 3 µl 1x-Ligase Puffer) gegeben, gevortext und in einem Thermocycler mit folgendem Programm inkubiert: 16 Stunden bei 4°C, 10 Minuten bei 65°C, Halten bei 10°C. Die Herstellung der Adapter erfolgte wie unter Kapitel 2.2.5 auf Seite 44 beschrieben, die verwendeten Oligonukleotide (gekennzeichnet mit „rapid“) sind dem Abschnitt 2.1.5 auf Seite 35 f zu entnehmen. Die erhaltenen Ligationsprodukte wurden zweimal mit SPRI-Beads wie in Kapitel 2.2.3 auf Seite 43 beschrieben bei einem Längen-Cutoff von 500 bp aufgereinigt. Am Ende wurde die DNA in 20 µl EB-Puffer eluiert, wobei das Eluat per Vakuum-Zentrifugation auf 6 µl eingengt wurde. Zur Kontrolle wurde die Library auf einen High-Sensitivity-Chip aufgetragen und analysiert. Die Ergebnisse der Konzentrationsbestimmungen und –profile sind auf Seite 116 f. dargestellt.

Größenunterschluss. Sollten auf dem Profil des High-Sensitivity-Chips unerwünschte Nebenprodukte, die größer als 800 bp sind, zu sehen sein, wurde die Library wie folgt aufgereinigt: die Library wurde auf ein Volumen von 50 µl mit EB-Puffer in einem 1,5 µl Eppi gebracht, und die Menge Ampurebeads dazugegeben, bei der in der Titration der 1000 bp-Peak nicht abgeschwächt wurde, und gleichzeitig der 700 bp-Peak kaum vorhanden gewesen ist. Die Suspension wurde gevortext, 5 min bei Raumtemperatur stehen gelassen, für 5 min im MPC separiert, und anschließend der Überstand in ein neues 1,5 ml Eppi übertragen. Dieser Überstand wurde mit Ampurebeads aufgefüllt, bis ein Gesamtvolumen von 100 µl erreicht ist. Die Suspension wurde wieder gevortext, 5 min bei Raumtemperatur inkubiert, für 5 min im MPC separiert, der Überstand abgezogen und verworfen, zweimal mit 70%igem Ethanol gewaschen, mit 20 µl EB-Puffer eluiert und das Eluat per Vakuum-Zentrifugation auf 6 µl eingengt; es wurde erneut eine Kontrolle auf einem High Sensitiv-Chip aufgetragen.

454-Runstart. Die so erhaltenen Librarys wurden in einer emPCR amplifiziert (Ansatz siehe Abb. 40) und mit den im Beadrecovery erhaltenen Beads die Region einer PTP beladen und ein Sequenzierlauf gestartet, wobei die unter Kapitel 2.2.1 und 2.2.2 beschriebenen Protokolle zur Anwendung kamen.

Library Name	tMID	Konzentra- tion in ng/μl	durchschnittl. Länge in bp	Konzentration in Moleküle/μl	Anteil der Beads in %	cpb	μl Library
RA-BS_03LK	29	0,02563	650	36163888,56	3,21	4,59	14,22
RA-BS_10LK	31	0,01611	790	18702872,87	2,78	3,94	20,48
RA-BS_02H	32	0,11441	773	135745163,4	21,37	4,69	25,81
RA-BS_07H	26	1,38817	808	1575691407	3,63	0,41	0,33
RA-BS_14LK	33	0,86453	803	987425630,1	1,50	0,19	0,99
Kareti LK	23	0,38929	607	588199200	0,43	0,49	1,24
RA-BS_12H	27	2,57035	820	2874870490	3,85	0,34	1,60
TM51H	24	0,416	762	500700738,6	12,82	0,21	1,84
ND07LK	25	0,09901	637	142554030,5	0,43	0,25	2,57

Abb. 40: emPCR-Ansatz mit den Librarys hergestellt nach verbessertem Protokoll

„Anteil der Beads in Prozent“ beschreibt den Anteil an Capture-Beads, die von einer Library besetzt werden sollen. Weil nur ein Cup der emPCR angegeben wird, ist ihre Summe 50%. Die cpb-Werte sind anhand von Titrationen und vorhergehender Runs bestimmt worden, bei denen die Librarys untergemischt worden sind.

Die Ergebnisse von Beadrecovery und des Sequencing-Runs werden auf Seite 116 ff. gezeigt.

3. Ergebnisse

In Kapitel 3.1 werden die Ergebnisse einer Ampurebead-Titration exemplarisch für eine Charge dargestellt. Ab Kapitel 3.2 werden die Sequenzierungsversuche von cDNAs auf dem 454 nach Standard-Protokollen beschrieben, deren Fehlschläge zu der Entwicklung einer neuen Methode für die gerichtete Sequenzierung auf dem 454 führten, deren Ergebnisse ab Kapitel 3.3 präsentiert werden. Eine inhaltliche Übersicht dieser Entwicklung liefert Abb. 41.

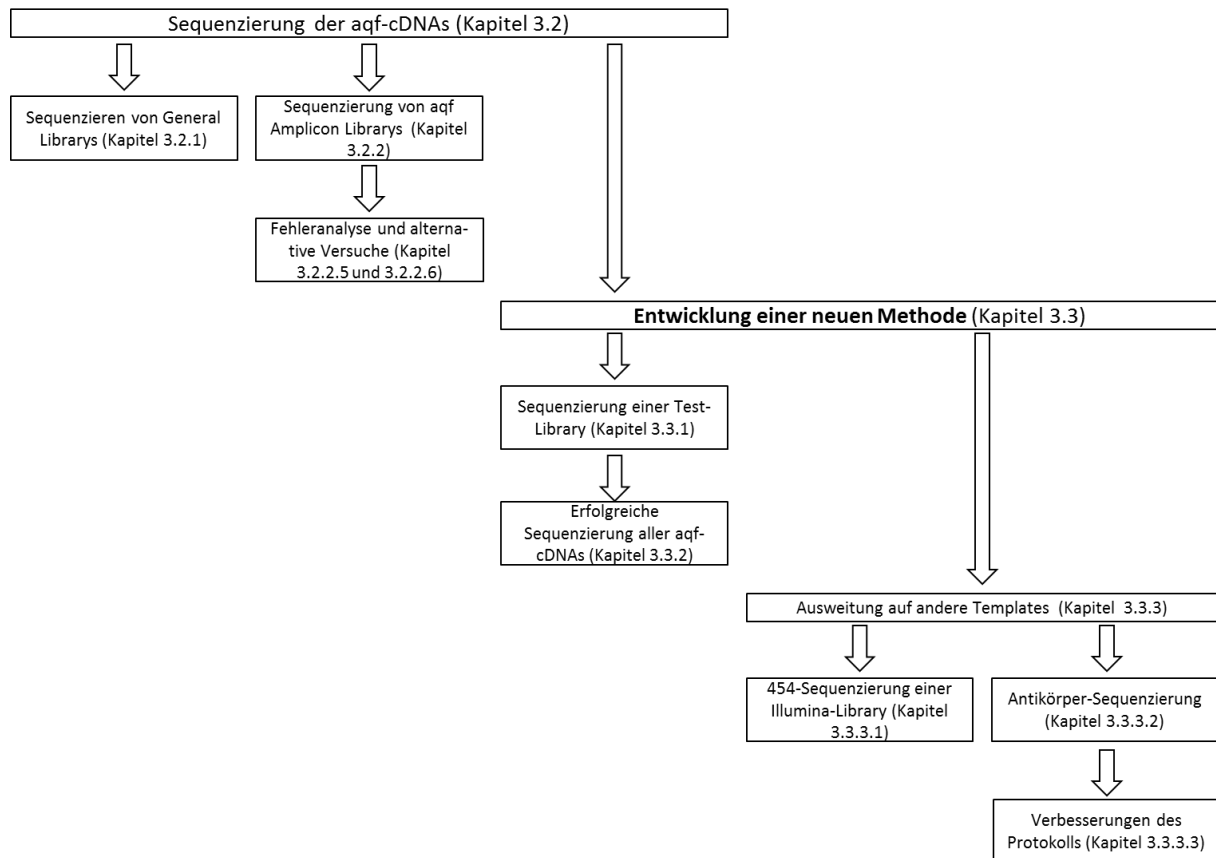


Abb. 41: Übersicht des Ergebnisteils

3.1 Ampurebead-Titration

Für die erfolgreiche 454-Sequenzierung ist eine Aufreinigung der DNA bei der Library-Herstellung essentiell. Das Entfernen zu kleiner DNA-Fragmente stellt dabei einen wichtigen, wenn nicht sogar den wichtigsten Schritt bei der 454-Library-Herstellung dar. Diese Aufreinigung geschieht mit Ampurebeads mittels SPRI-Methodik, da sie nicht nur sehr effizient ist, sondern auch die Steuerung der zu entfernenden Fragmentgrößen ermöglicht.

Um das beste DNA-SPRI-Bead-Verhältnis jeder Charge für einen idealen Größencutoff, d.h. Aufreinigung und Ausschluss von kleineren Fragmenten, zu ermitteln (Seite 44), wurde Roche's DNA

Molecular Weight Marker XIV mit verschiedenen Ampurebead-Mengen aufgereinigt. Es werden als Beispiel die Ergebnisse einer Charge präsentiert (Abb. 42 und Abb. 43, Seite 66 und 67). Die Peak-Ratios werden anhand der Ergebnisse des DNA-7500-Chips berechnet (Tab. 19).

Volumen- Verhältnis	Konzentrationen in ng/μl					Peak-Ratio (Konzentrations-Quotient)			
	200 bp	300 bp	400 bp	500 bp	900 bp	200/900	300/900	400/900	500/900
50 : 100	0,00	0,00	0,00	0,43	0,79	0,00	0,00	0,00	0,54
55 : 100	0,00	0,00	0,33	2,03	1,32	0,00	0,00	0,25	1,54
60 : 100	0,00	0,29	0,67	2,87	1,07	0,00	0,27	0,63	2,68
65 : 100	0,00	0,74	1,21	3,47	1,16	0,00	0,64	1,04	2,99
70 : 100	0,00	0,71	1,00	2,63	0,83	0,00	0,86	1,20	3,17
75 : 100	0,47	1,26	1,58	3,80	1,28	0,37	0,98	1,23	2,97
80 : 100	0,40	0,89	1,09	2,78	0,92	0,43	0,97	1,18	3,02
85 : 100	0,59	1,06	1,29	3,15	1,02	0,58	1,04	1,26	3,09
90 : 100	1,10	1,87	2,24	5,49	1,87	0,59	1,00	1,20	2,94
95 : 100	0,62	1,01	1,25	3,10	1,03	0,60	0,98	1,21	3,01
100 : 100	1,05	1,66	2,01	4,91	1,65	0,64	1,01	1,22	2,98
Kontrolle	1,96	3,04	3,69	8,56	3,08	0,64	0,99	1,20	2,78
Vorgaben von Roche für 454-Librarys:						0,00	0,40	0,70	2,90

Tab. 19: Berechnung der Peak-Ratios bei der Ampurebead-Titration

Ab einem Volumenverhältnis von 65:100 verändert sich der Konzentrations-Quotient von 500/900 nicht mehr (der Quotient der Kontrolle ist mit 2,78 sogar etwas kleiner), d.h. bei Volumina von 65 μl Ampurebeads aufwärts bezogen auf 100 μl DNA werden so gut wie alle 500 bp Fragmente an die Beads getrieben und gehen nicht im Überstand verloren. Beim Volumen-Verhältnis von 100:100 werden auch die 200 bp Fragmente vollständig an die Beads getrieben (die Peak-Ratio 200/900 entspricht der der Kontrolle). Vergleicht man die DNA 7500-Chip-Messung vom Volumen-Verhältnis 100:100 mit der Messung der Kontrolle (Abb. 43 auf Seite 67), so wird ersichtlich, dass die 100 bp Fragmente des Markers beim Verhältnis von 100:100 vollständig weggereinigt wurden, obwohl die 200 bp-Fragmente vollständig erhalten blieben. Die Trennung in diesem Größenbereich erfolgt also sehr strikt, weshalb die Ampurebeads generell hervorragend für die Aufreinigung von PCR-Ansätzen zum Entfernen von Nukleotiden und Primern geeignet sind.

Die Vorgaben von Roche für die Peak-Ratios, die als ideal für die Aufreinigung von Librarys von zu kleinen Fragmenten angesehen werden, liegen bei Null für 200/900, bei 0,4 für 300/900, bei 0,7 für 400/900 und bei 2,9 für 500/900. Vergleicht man dies mit den berechneten Peak-Ratios (Tab. 19), so liegen die Vorgaben für Roche zwischen den Peak-Ratios der Volumenverhältnisse von 60:100 und 65:100 also dem Einsatz zwischen 60 bis 65 μl Ampurebeads auf 100 μl DNA.

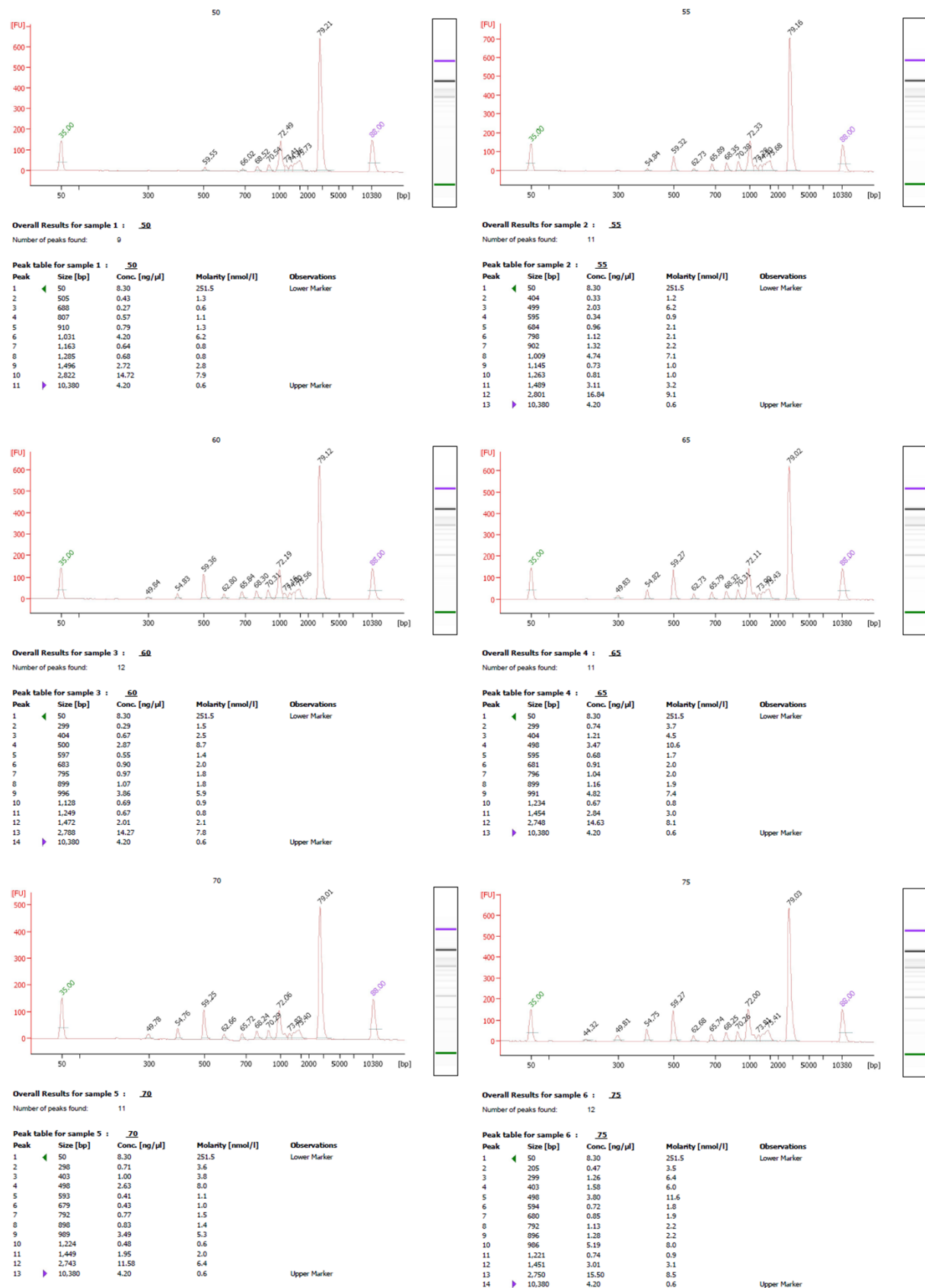


Abb. 42: Ampurebead-Titration auf dem DNA 7500 Chip

Es sind die Ergebnisse der Titrationen von 100 μl verdünntem DNA Molecular Weight Marker XIV mit 50 bis 75 μl Ampurebeads dargestellt.

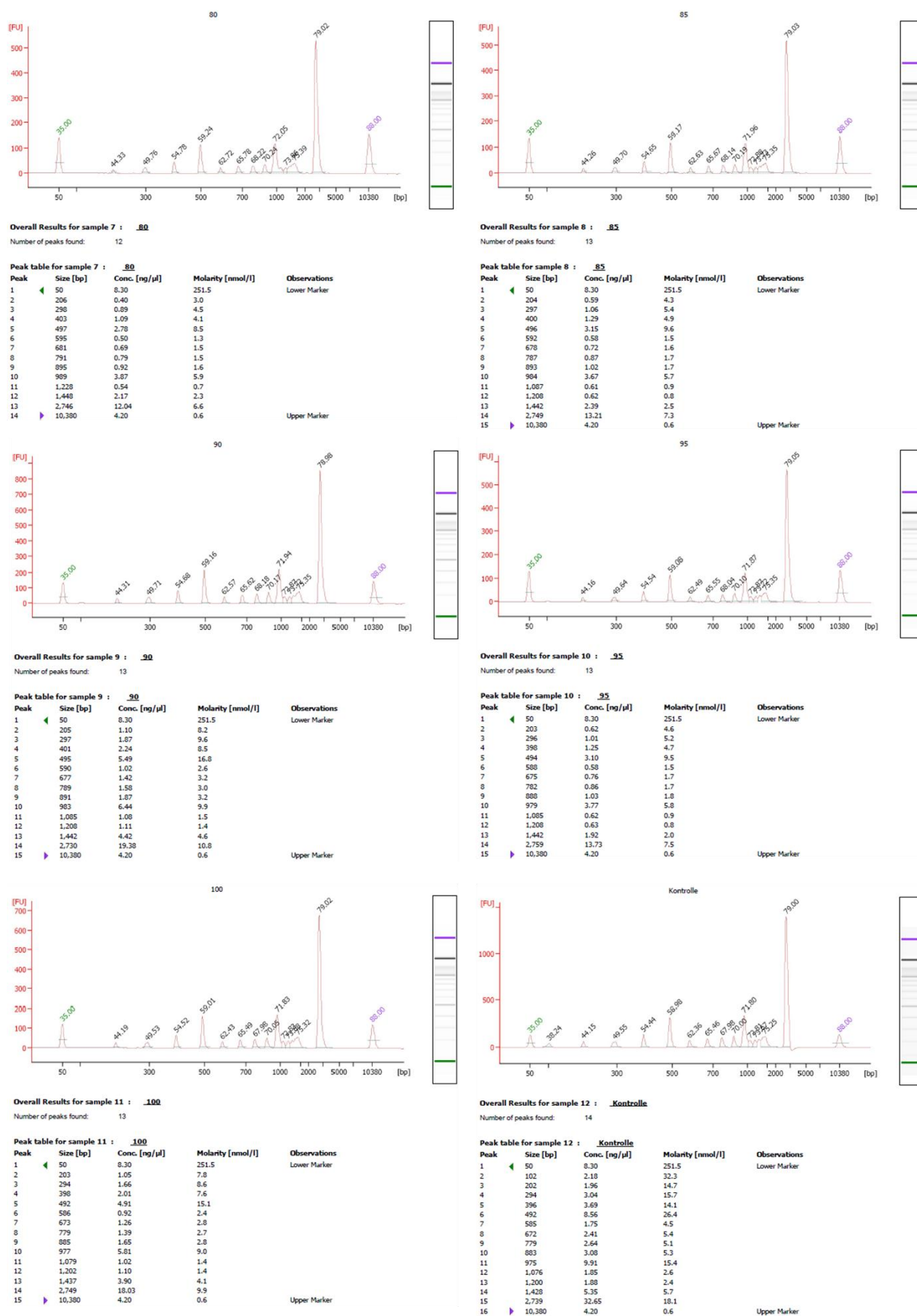


Abb. 43: Fortsetzung der Ergebnisse der Ampurebead-Titration auf dem DNA 7500 Chip

Es sind die Ergebnisse der Titrationen von 100 μl verdünntem DNA Molecular Weight Marker XIV mit 80 bis 100 μl Ampurebeads sowie der Kontrolle dargestellt.

3.2 Sequenzierungsversuche des aqf-Projektes

Zur Sequenzierung der aqf-cDNAs musste aus diesen 454-Librarys hergestellt werden. Es wurden zwei unterschiedliche Strategien angewandt: die Sequenzierung der cDNAs über a) die Erstellung von Titanium General Librarys und b) die Erstellung von Titanium Amplicon Librarys.

3.2.1 Sequenzierung mittels Titanium General Library-Methodik

Aus vier aqf-cDNAs wurden Titanium General Librarys hergestellt, diese titriert und anschließend nach emPCR auf dem FLX Instrument sequenziert (vergleiche Seite 45 ff.). Anhand der Run-Ergebnisse könnte man dann erkennen, ob der 454-Lauf erfolgreich war und sich die General-Library-Methode, d.h. die Erstellung ungerichteter 454-Librarys über Random-Adapter-Ligation, für die Sequenzierung der aqf-cDNAs eignet.

3.2.1.1 Erstellung der Titanium General Librarys

Die aqf-cDNAs aqf3b, aqf4b, aqf5b und aqf6b wurden mittels Ultraschall fragmentiert, über SPRI aufgereinigt um Fragmente kleiner 500 bp loszuwerden und je ein Aliquot auf einem 7500 DNA-Chip aufgetragen und analysiert (Abb. 44).

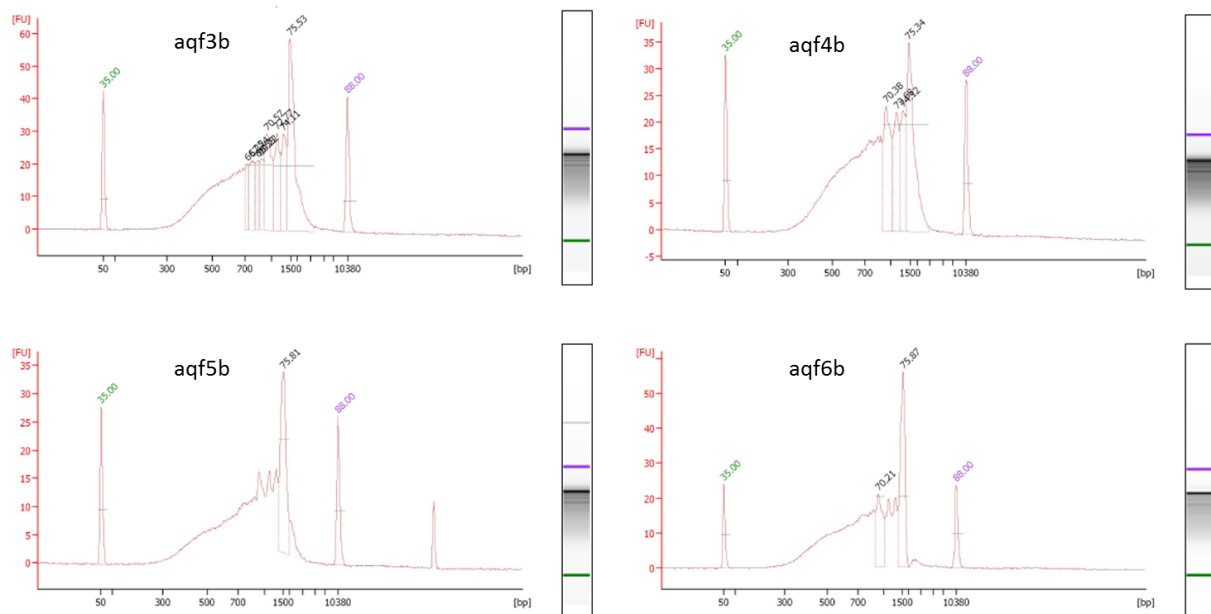


Abb. 44: Ergebnisse der aqf-cDNAs auf dem 7500 DNA-Chip

Alle vier aqf-cDNAs liegen in einem Größenbereich von 300 bp bis 1500 bp (Abb. 44), wobei die Konzentration der Fragmente zu einem deutlich herausstechenden Peak bei 1500 bp hin ansteigt. Teilweise lassen sich auch noch einzelne Peaks zwischen 700 bp und 1500 bp erkennen. Auf der einen Seite wurden die vier cDNA-Librarys fragmentiert, da ein „DNA-Schmier“ vorhanden ist, auf der anderen Seite war die Fragmentierung nicht vollständig, da Peaks erkennbar sind bzw. ein Peak bei 1500 bp deutlich heraussteicht. Vergleicht man die Größe bzw. Höhe der Flächen unterhalb der FU-

Messkurve der einzelnen cDNAs, so fällt auf, dass diese bei allen cDNAs annähernd gleich groß ist, was bedeutet, dass die vier cDNAs, was die Fragmentierungsprodukte angeht, in annähernd gleichen Mengen vorliegen. Im Vergleich zu der Höhe der Fragmentierungsprodukte der vier cDNAs ist die Höhe des Peaks bei 1500 bp unterschiedlich groß. Beides kann mit einer nicht vollständigen Fragmentierung erklärt werden; 1 min Fragmentierungszeit ist lang genug, um entsprechende Mengen an fragmentierter cDNA zu generieren, reicht aber nicht für eine vollständige Fragmentierung der gesamten cDNA aus. Bei keiner der vier cDNAs sind Fragmente kleiner als 300 bp erkennbar; und Fragmente kleiner als 500 bp wurden deutlich reduziert, was an einem stärkeren Abfall der Fluoreszenz-Units (FU) zwischen 500 bp und 300 bp deutlich wird. Dieser Verlauf, also das Fehlen von Fragmenten kleiner als 300 bp und eine starke Aufreinigung von Fragmenten kleiner 500 bp, entspricht den Vorgaben von Roche, die als ideal angesehen werden (vergleiche Peak-Ratios in Tab. 19 auf Seite 65).

Die fragmentierten und SPRI-aufgereinigten cDNAs wurden dem General-Library-Protokoll entsprechend weiter verarbeitet. Es folgten ein Endrepair der fragmentierten-DNA, die Ligation von selbst hergestellten 454-Adapter mit MIDs an die cDNAs, eine weitere SPRI-Aufreinigung zum Entfernen von ungebundenem Adapter und einer weiteren, besseren Größenselektion, ein Nick-Repair an magnetischen Beads, die aufgrund der Adapterstruktur nötig ist, und als letzter Schritt die Elution von einzelsträngiger 454-Library-cDNA von den magnetischen Beads. Diese einzelsträngigen 454-Librarys werden zur Kontrolle auf einen RNA Pico 6000 Chip aufgetragen (Abb. 45 auf Seite 70).

Die vier aqf-cDNAs sind am Ende der Library-Präparation in deutlichen Mengen vorhanden. Die Messung von aqf6b weicht von den anderen ab; es gibt viele Ausschläge entlang der Messlinie; auch eine Konzentrations-Bestimmung über den RNA-Chip war deshalb nicht möglich. Die Konzentration der aqf6b-Library wurde mittels Qubit-Messung auf 69 pg/µl bestimmt. Kurvenverläufe wie bei aqf6b sind meistens auf eine verschmutzte Messelektrode, Luftblasen in der Probe, zu hohem Salzgehalt oder sonstige Kontaminationen zurückzuführen. Es waren keine Luftblasen sichtbar, normalerweise gibt der Agilent Bioanalyzer auch eine Fehlermeldung heraus („Voltage Error“) und bricht die Messung zu Beginn ab beim Vorhandensein von Luftblasen; letztendlich können sie aber auch nicht ausgeschlossen werden. Eine verschmutzte Messelektrode kann dagegen ausgeschlossen werden; ist die Elektrode verschmutzt, so betrifft das selten nur eine Position auf dem Chip; alle anderen gemessenen Positionen (cDNAs, Leer-Probe und Ladder) sind davon aber nicht betroffen. Außerdem zeigten folgende Messungen an späteren Tagen keine Auffälligkeiten auf der Elektroden-Position von aqf6b (Ergebnisse nicht gezeigt), so dass auch die Verunreinigung eines einzelnen Pins der Messelektrode ausgeschlossen werden kann. Bei zu hohem Salzgehalt in Proben können auch solche starken Ausschläge beobachtet werden; ob das hier bei aqf6b die Ursache dafür ist, kann nicht ausgeschlossen werden, da es bei der Library-Präparation zum „Mitschleppen“ von salzhaltigen Puffern (z.B. Neutralization-Solution) kommen kann durch nicht vollständig ausgeführte Waschschriffe oder dem Übersehen und Verbleiben von Puffer-Spritzern an den Wänden der verwendeten Tubes.

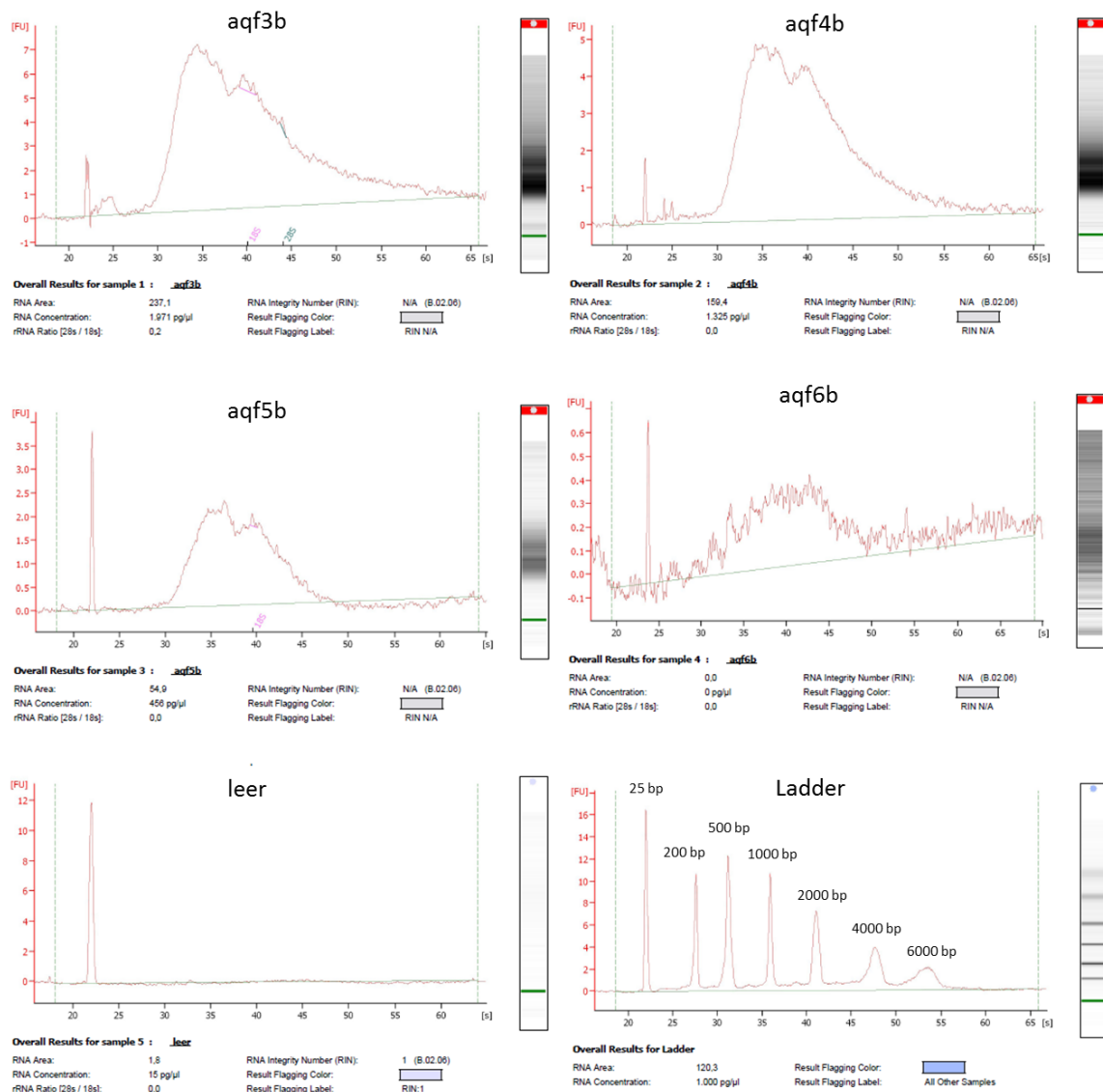


Abb. 45: aqf-Titanium General Libraries auf dem RNA Pico 6000 Chip

Dargestellt sind die vier cDNA-Librarys, eine Leerprobe (leer) und der im RNA Pico 6000 Chip-Kit enthaltene Marker (Ladder), der bei jedem Chip mitgemessen wird. Als Vergleich sind über die Marker-Banden die entsprechenden Größen aufgeführt. Anhand der Agilent Messung des RNA-Chips lassen sich folgende Konzentrationen ablesen: aqf3b hat 1,971 ng/μl, aqf4b hat 1,325 ng/μl, aqf5b 0,456 ng/μl.

Aus welchem Grund auch immer die starken Ausschläge entlang der Messkurve bei aqf6b zustande kommen, so ist auch bei aqf6b erkennbar, dass die Messkurve im Prinzip einen ähnlichen Verlauf hat wie die der drei anderen cDNAs: sie startet bei 30 s und fällt schließlich zu 45 s stark ab. Bei den anderen drei cDNAs bewegt sich der Großteil der gemessenen DNA in einem Bereich zw. 30 und 40 s. Über die Ladder kann man den gemessenen Zeiten die entsprechenden Basenlängen entgegensetzen: der Bereich zw. 30 und 40 s entspricht DNA-Fragmentlängen von knapp unterhalb 500 bp bis unter 2000 bp, vermutlich 1500 bp.

Vergleicht man die Messkurvenverläufe des RNA Pico 6000 Chips mit denen des 7500 DNA-Chips, so fällt auf, dass der beim DNA-Chip stark hervortretende Peak bei 1500 bp nicht mehr so deutlich beim RNA-Chip zu sehen ist, wenn er auch dort auftritt (bei rund 40 s). Außerdem scheinen die Library-Fragmente, welche kleiner als 500 bp sind und auf dem DNA-Chip noch deutlich zu sehen waren (der

Bereich zwischen 300 bp und 500 bp) jetzt zu fehlen, was mit der zweiten SPRI-Aufreinigung nach der Adapter-Ligation zu erklären ist. Auf der anderen Seite kann man bei aqf3b und aqf4b zwei kleine Peaks unterhalb von 25 s auf dem RNA-Chip erkennen. Bei aqf5b fehlen diese und bei aqf6b ist keine Aussage darüber möglich, aufgrund der schlechten Messung. Da auf dem DNA-Chip in dem Bereich keine Peaks vorhanden waren, sind es entweder Fehlmessungen oder es handelt sich um echte DNA, die bei der Library-Präparation in einem späteren Schritt dazugekommen ist. Eine Fehlmessung erscheint unwahrscheinlich, da die beiden Peaks unterhalb 20 s bei aqf3b und aqf4b die gleiche Länge zu haben scheinen. Wenn es sich nicht um irgendwelche zufälligen Kontaminationen handelt, die in die Präparationen von aqf3b und aqf4b gelangt sind (was auch unwahrscheinlich ist, da die gesamte Library-Herstellung unter einer vorher desinfizierten Clean-Bench stattfand), dann können diese beiden Peaks Adapter-Reste darstellen. Adapter A mit einem 10 bp MID hat eine Länge von 40 bp und Adapter B ist 30 bp lang; mögliche Dimere daraus dann 80, 70 oder 60 bp, was in den Bereich um und kurz unterhalb 25 s auf dem RNA-Chip fällt (vergleiche mit den Basenangaben der Ladder).

Aus den Kurvenverläufen von DNA- und RNA-Chip muss die durchschnittliche Länge jeder Library abgeschätzt werden, da diese für die Berechnung der Library-Konzentration in Molekülen pro Volumen benötigt wird (für den Ansatz der emPCRs). Zu beachten bei dieser Schätzung ist, dass für eine Vorhersage für die emPCR und damit verbunden guten Enrichment-Rates in Betracht gezogen werden muss, dass sich kleinere Moleküle in der emPCR besser amplifizieren lassen als größere. Würde man diese Tatsache außen vor lassen, so könnte man die durchschnittliche Länge der vier aqf-Librarys anhand des DNA-Chips auf ca. 900 bis 1000 bp schätzen, da die Messkurven sich nach rechts neigen. Bezieht man die gerade geführten Überlegungen in die Schätzung mit ein, so müsste die durchschnittliche Länge bei weniger als 900 bis 1000 bp liegen; daher wird die durchschnittliche Länge der vier aqf-General Librarys auf 700 bp geschätzt.

3.2.1.2 Titration der General Librarys

Die vier General Librarys aus aqf-cDNA wurden titriert (Tab. 20 und Abb. 46, Seite 72 und 73), wobei jeweils cpb-Zahlen von 0,5, 1, 2 und 4 zur Anwendung kamen, um eine ideale cpb-Zahl zu bestimmen, mit der die Librarys dann in einen großen emPCR-Ansatz sequenziert werden können.

Allen Librarys ist gemeinsam, dass mit steigenden cpb-Zahlen die Enrichment-Rates ebenfalls ansteigen, was auch nicht anders zu erwarten war: erhöht man die DNA-Konzentration, so sollte auch die Zahl der DNA-tragenden Beads steigen. Die Beadrecovery-Rates liegen in einem Bereich von 65 bis kurz über 100 %. Das bedeutet auf der einen Seite, dass die Rückgewinnung an Beads aus der Emulsion erfolgreich war, von Roche werden Beadrecovery-Rates über 65% über Spritzen als „typisch“ bezeichnet [122], auf der anderen Seite zeigen Werte über 100% (Beadrecovery-Rate von aqf5b bei 2 cpb), dass entweder die Messung mit dem Partikelzähler fehlerhaft ist (Pipettierfehler, Luftblasen in der Messung), oder bei der Vorbereitung der Titration zu viele Capture-Beads pipettiert wurden. Über die Bestimmung der Enrichment-Rates soll eine für jede Library gute cpb-Zahl gewählt werden können. Die Obergrenze dabei wird von Roche zwischen 15-20% angegeben. Schaut man sich die errechneten Enrichment-Rates an, so liegt nur aqf5b unterhalb von 15% bei 0,5 und 1 cpb.

Bei 0,5 cpb hatte aqf5b 7,9% angereicherte Beads. Da dieser Wert ausreichend ist, um genügend Beads selbst für eine ganze Region zu erzeugen, werden 0,5 cpb für aqf5b bei dem folgenden Sequencing-Run verwendet (bei einer angenommenen Beadrecovery-Rate von 85% - „typisch“ laut Roche für große Ansätze [74], erhält man bei einer Enrichment-Rate von 7,9% für eine Region 2,35 Mio. Beads; 2,0 Mio. Beads werden pro Region für das Sequencing benötigt).

Library	cpb	zurückgewonnene Beads	Beadrecovery-Rates in Prozent	aufgeklärte Beads	Enrichment-Rates in Prozent
aqf3b	0,5	717600	65,8	124960	17,4
	1	861600	79,0	187680	21,8
	2	1054500	96,7	288860	27,4
	4	836400	76,7	418600	50,0
aqf4b	0,5	715650	65,7	142170	19,9
	1	659600	60,5	132090	20,0
	2	1004550	92,2	275220	27,4
	4	1065300	97,7	378180	35,5
aqf5b	0,5	962200	88,3	75680	7,9
	1	892000	81,8	119890	13,4
	2	1093500	100,3	169200	15,5
	4	823400	75,5	267740	32,5
aqf6b	0,5	976500	89,6	428000	43,8
	1	1059850	97,2	574420	54,2
	2	971550	89,1	715440	73,6
	4	1040600	95,5	725190	69,7

Tab. 20: Titration der General Librarys

Die Tabelle zeigt die mittels Partikelzählers gewonnenen Messergebnisse der Beads nach Beadrecovery (zurückgewonnene Beads) und Enrichment (aufgeklärte Beads) und die sich daraus ergebenden prozentualen Recovery-Rates. Für jede cpb-Zahl wurde bei der Beadmessung eine Dreifachbestimmung durchgeführt; in der Tabelle sind die entsprechenden Mittelwerte angegeben.

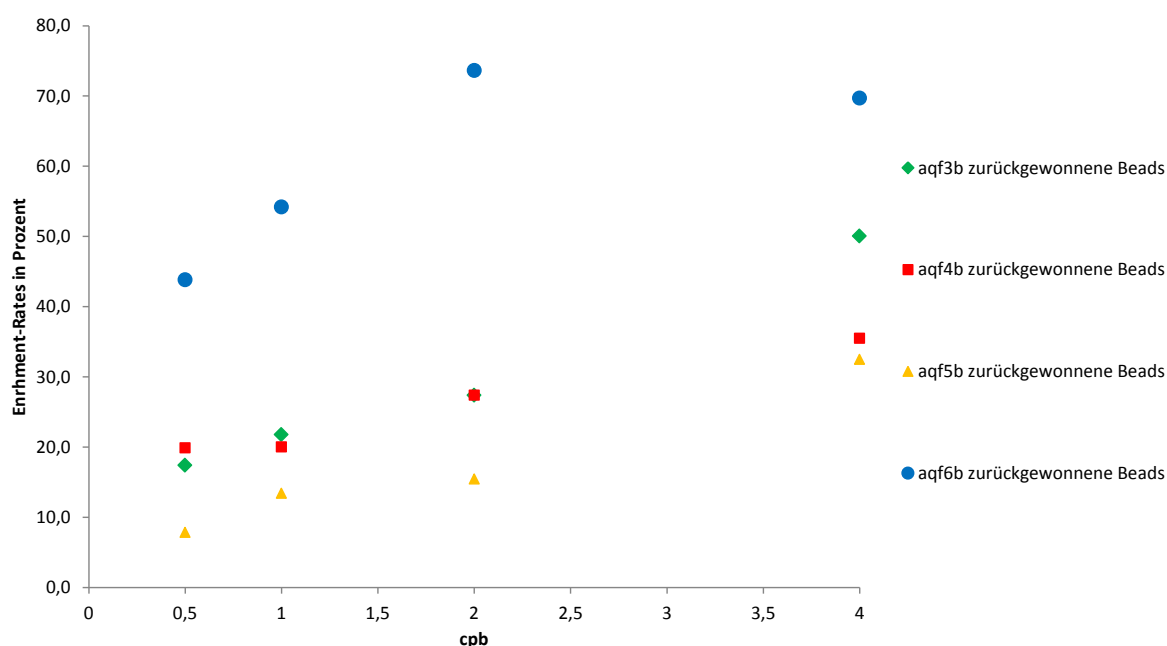


Abb. 46: Graphische Darstellung der Titrationsergebnisse

Bei aqf3b, aqf4b und aqf6b liegen die Enrichment-Rates über 15% bei allen cpb-Zahlen. Die cpb-Zahlen, die für diese Librarys später verwendet werden sollen, müssen also geschätzt werden. Für die Schätzung wird im Anfangsbereich ein linearer Abfall zu einer Enrichment-Rate von Null bei Null cpb angenommen und die Messkurven graphisch ausgewertet. Bei einer zu erreichenden Enrichment-Rate von 8% werden so für aqf3b 0,4 cpb, für aqf4b 0,35 cpb und für aqf6b 0,05 cpb abgelesen und diese für den Ansatz der emPCR verwendet (siehe Seite 48).

3.2.1.3 Beadrecovery der General Librarys

Mit den vier aqf-General Librarys wurde eine große emPCR mit anschließendem Beadrecovery (Tab. 21) für einen Sequencing-Runstart durchgeführt.

	Librarys	zurückgewonnene Beads	Beadrecovery-Rates in Prozent	aufgereinigte Beads	Enrichment-Rates in Prozent
Cup A	aqf3b aqf4b	29368500	83,9	4080705	13,9
Cup B	aqf5b aqf6b	28504000	81,4	608190	2,1

Tab. 21: Beadrecovery der emPCR von den General Librarys

Es werden die mittels Partikelzählers gewonnenen Messergebnisse der Beads von den beiden Cups nach Beadrecovery (zurückgewonnene Beads) und Enrichment (aufgereinigte Beads) gezeigt und die sich daraus ergebenden prozentualen Recovery-Rates.

Wie aus der Tabelle ersichtlich ist, wurden die Beads von beiden Cups mit einer Ausbeute von über 80% aus den Emulsionen zurückgewonnen. Das liegt nur knapp unterhalb der von Roche angegebenen Rate von 85% [85] -das Rückgewinnen der Beads war erfolgreich. Die Enrichment-

Rates beider Cups liegen unter dem von Roche vorgegebenen Grenzbereich von 15-20%, d.h. die Mehrzahl der Beads sollte monoklonal und gut sequenzierbar sein. Mit 2,1% Enrichment-Rate liegt Cup B deutlich unterhalb der Erwartungen; die Librarys in Cup B sollten eigentlich auf eine Sequencing-Region geladen werden, wozu es mindestens zwei Mio. Beads bräuchte. Cup A hat dagegen mehr als genug Beads für eine Region geliefert; insgesamt hat das Beadrecovery rund 4,7 Mio. Beads gebracht, was genug für einen Sequenzierlaufstart ist. Auffällig bei den Enrichment-Rates beider Cups ist die Tatsache, dass sie von den Erwartungen der Titration abweichen; bei den aus der Titration angewandten cpb-Werten wurden Enrichment-Rates im Bereich von 7-10% bei beiden Cups erwartet; besonders Cup B liegt deutlich unterhalb dieser Marke.

3.2.1.4 Sequencing-Run der General Librarys

Die in der emPCR amplifizierten und im Beadrecovery aufgereinigten Beads, die die vier aqf-General Librarys tragen, wurden auf eine zwei Regionen-PTP geladen und dann auf dem 454 sequenziert. Tab. 22 zeigt die Ergebnisse des Runs, in den Abbildungen auf der folgenden Seite sind sie graphisch dargestellt; Tab. 23 und Abb. 50 zeigt eine Analyse der Kontroll-DNA.

	TCAG-Wells (Library)			CATG-Wells (Control)		
	Region 1	Region 2	Total	Region 1	Region 2	Total
Raw Wells	794389	890689	1685078	794389	890689	1685078
Key Pass Wells	764115	858483	1622598	12212	8504	20716
Passed Filter Wells	278354	200513	478867	9911	6984	16895
Total Bases	78080261	55975732	134055993	3043471	2087181	5130652
% Passed Filter	36,43	23,36	29,51	81,16	82,13	81,56
Lenght Average	280,51	279,16	279,94	307,08	298,85	302,68
Lenght Std Deviation	161,66	150,27		132,85	126,99	
Longest Reads Lenght	1191	1196	1196	849	541	849
Shortest Reads Lenght	40	40	40	40	40	40
Median Reads Lenght	322	308	315	312	305	308
Failed Dot	49129	16510	65639	79	11	90
Failed Mixed	251415	476481	727896	176	191	367
Failed Short Quality	176967	158629	335596	1881	1296	3177
Failed Short Primer	4849	2269	7118	0	0	0
% Dot + Mixed	39,33	57,43	48,91	2,09	2,38	2,21
% Short	23,79	18,74	21,12	15,40	15,24	15,34

Tab. 22: Run mit den aqf-General Library

Es sind die Ergebnisse des Runs dargestellt, die von der GSRUNBrowser-Software ausgegeben werden. Die Ergebnisse sind nach Keys geordnet; links befinden sich die der Library-DNA (tcag-wells) und rechts die der beim Sequencing mitgeladenen Kontroll-DNA (catg-wells).

Bei diesem Sequenzier-Run kam eine Gesamtmenge von nur 478867 Reads herunter (Passed Filter Wells, Tab. 22), das sind deutlich weniger als die von Roche angegebene Untergrenze von 800000 Reads für einen guten Lauf. Der Anteil der ausgegebenen Reads bezogen auf die Key Pass Wells,

also aller Wells, die einen TCAG-Key besitzen, ist mit 29,51% niedrig (% Passed Filter Wells). Die durchschnittliche Länge liegt bei rund 280 bp (Length Average, Total), was ebenfalls deutlich unter den Vorgaben von 350-400 bp für einen guten Lauf liegt. Die Größenverteilung der Reads auf beiden Regionen sieht entsprechend aus (Abb. 47 auf Seite 76); insbesondere im Bereich zwischen 40 und 100 bp scheint es viele Reads zu geben. Dieser Sequenzier-Lauf ist schlecht.

Schaut man sich den Anteil der Reads an, die beim Signal-Processing ausgefiltert wurden (Failed Reads), so stellt man fest, dass dieser bei beiden Regionen sehr hoch ist; es wurden rund 70% der TCAG-Key tragenden Reads, also potentieller Library-DNA, verworfen (rund 49% Dot + Mixed und 21% Short). Region 2 scheint dabei insgesamt noch etwas schlechter zu sein als Region 1 (vergleiche auch Abb. 49 auf Seite 77), der Anteil der Passed Filter Wells liegt mit 23% noch deutlich unter Region 1 mit 36%. Bei den ausgefilterten Reads fällt auf, dass der weitaus größte Anteil vom Mixed Filter entfernt wurden („Failed Mixed“, Abb. 49 auf Seite 77); auf Region 1 waren das 251415 Reads und auf Region 2 sogar 476481 (Failed Mixed, TCAG-Wells, Tab. 22 auf Seite 74). Der Anteil an Reads, die als „Short“ aussortiert wurden, liegt mit rund 21% im üblichen Bereich bzw. ist nicht auffällig.

Die Anzahl an Raw-Wells von Library-Beads liegt mit rund 800000 auf Region 1 und rund 900000 auf Region 2 nur knapp unterhalb den Vorgaben von Roche (1-1,2 Mio. Raw-Wells pro Region).

Von der Kontroll-DNA haben auf beiden Regionen rund 82% (Passed Filter Wells) der CATG-Key-tragenden Sequenzen das Filtering im Signal-Processing passiert (vergleiche Tab. 22 auf Seite 74 und Abb. 49 auf Seite 77) -bei sehr guten Runs werden Werte um die 90% erwartet. Die durchschnittliche Länge der Kontroll-DNA-Reads liegt bei rund 303 bp (Length Average, CATG-Wells, Total); bei einem guten und fehlerfreien Lauf sollte sie zwischen 370 und 400 bp liegen, sie ist also deutlich kürzer. Das spiegelt sich dementsprechend auch in der Größenverteilung der Kontroll-DNA wieder (Abb. 48 auf Seite 76), der Anteil der kürzeren Reads ist etwas höher und der Großteil der Failed Reads wurde „Failed Short Quality“ herausortiert (Abb. 49 auf Seite 77, CATG-Wells). Im Gegensatz zu den Library-Reads gibt es bei den Kontroll-DNA-Reads keinen merklichen Unterschied zwischen Region 1 und 2; die prozentuale Anzahl an Passed Filter Wells, Failed Wells und die durchschnittliche Länge sind auf beiden Regionen annähernd gleich groß.

Für die Beurteilung der Qualität der Kontroll-DNA-Reads spielt nicht nur die Filter-Werte und die Länge eine Rolle, sondern auch die Sequenzgenauigkeit der erhaltenen Kontroll-DNA-Reads (Tab. 23 und Abb. 50 auf Seite 77). Eine Übereinstimmung (mit richtigen Kontroll-DNA-Sequenzen) von 95% über eine Länge von 100 bp erreichen auf beiden Regionen rund 79% der Kontroll-DNA-Reads; normalerweise wird bei guten Läufen ein Wert von knapp über 90% erreicht. Der Anteil der Reads, die mit den Sequenzen übereinstimmen, fällt ab, umso länger die geprüfte Sequenz wird. Das ist auf der einen Seite normal, da das Pyrosequencing ungenauer wird, umso länger die Reads werden bzw. die Sequencing-Flows voranschreiten, auf der anderen Seite ist der Abfall bei diesem Run doch stärker als erwartet; insbesondere der Anteil der Kontroll-Reads, die 95% oder 98% Übereinstimmung erreichen, ist bei allen Längen zu gering.

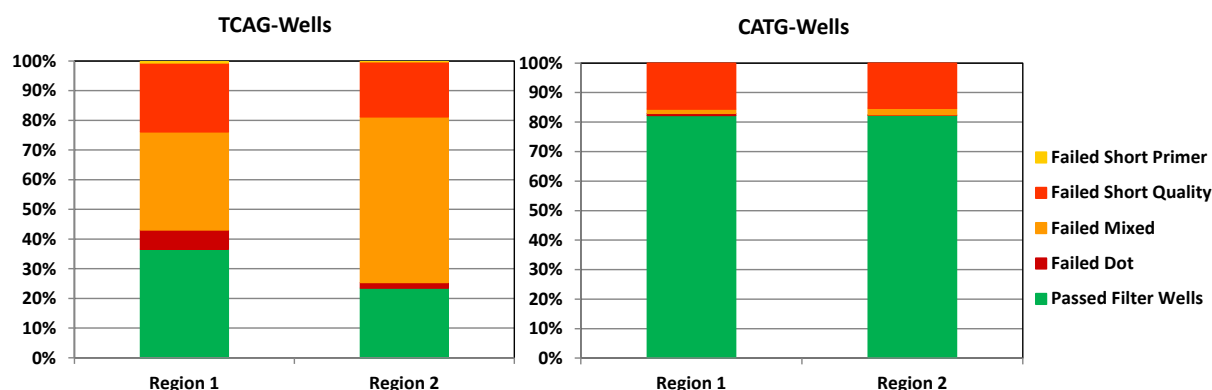


Abb. 49: Prozentuale Verteilung beim Run mit den aqf-General-Librarys

Es wird das Verhältnis auf beiden Regionen von Passed Filter Wells (grün) zu den Failed Wells dargestellt, angezeigt für Library-Reads (tcag-wells) und Kontroll-DNA-Reads (catg-wells).

Länge	100 bp			200 bp			300 bp			400 bp		
Match	100%	98%	95%	100%	98%	95%	100%	98%	95%	100%	98%	95%
Region 1	56,83	72,17	78,69	42,07	60,06	67,50	18,27	49,49	58,73	8,13	35,20	44,87
Region 2	57,28	72,60	79,20	43,52	61,48	69,58	18,19	49,19	59,21	6,61	32,37	43,39

Tab. 23: Kontroll-DNA-Analyse beim Run mit den aqf-General-Librarys

Es sind die von der GSRUNBrowser-Software ausgegebenen Ergebnisse der Kontroll-DNA-Analyse dargestellt. Für beide Regionen des Runs wird der prozentuale Anteil der Kontroll-Bead-Reads angegeben, der zu je 100%, 98% und 95% mit den Kontroll-DNA-Sequenzen übereinstimmt (Match). Diese Übereinstimmungen werden für Längen von 100 bp, 200 bp, 300 bp und 400 bp angegeben.



Abb. 50: Graphische Darstellung der Kontroll-DNA-Analyse

Es werden die Ergebnisse der Kontroll-DNA-Analyse aus Tab. 23 graphisch dargestellt.

Die schlechten Ergebnisse des Runs, die geringen durchschnittlichen Leseweiten und geringen Readausbeuten zeigen also, dass eine Sequenzierung der aqf-cDNA über die Erstellung von General-Librarys nicht erfolgreich war. In dem folgenden Abschnitt wird mit der einzigen Alternative, der Erstellung von Amplicon Librarys, versucht, eine Sequenzierung durchzuführen.

3.2.2 Sequenzierung mittels Amplicon-Methodik

Wie im vorangehenden Kapitel beschrieben, scheiterte die Sequenzierung der aqf-cDNAs mittels General-Library-Methodik. Als Alternative besteht die Möglichkeit der Herstellung einer Amplicon Library, die ein gerichtetes Sequenzieren ermöglicht. Aus sechs aqf-cDNAs wurden Amplicon Libraries erstellt. Dazu wurden die 454-Sequenzen mittels PCR an die cDNAs heranamplifiziert, die so erhaltenen 454-Libraries titriert und dann nach emPCR auf dem FLX Instrument sequenziert (vergleiche Seite 49 ff.).

3.2.2.1 Erstellung der Amplicon Libraries

Mit aqf3b, aqf4b, aqf5b, aqf6b, aqf9b und aqf10b wurden Amplicon Libraries hergestellt. Dazu wurden bei jeder Library die zwei verschiedenen Reverse-Primer „ADA CDNA“ und „ADA TANKER“ verwendet; die Forwardprimer unterscheiden sich nur in ihrer MID-Sequenz. Die Amplicons wurden in einer Gradienten-PCR generiert und Aliquots der Amplifikate auf ein Gel aufgetragen (Abb. 51, Abb. 52 und Abb. 53 auf den folgenden Seiten). Mit „M“ sind die Marker-Lanes gekennzeichnet (4kb, 2kb, 1kb und 0,5kb-Banden; Eigenproduktion des MPI für molekulare Genetik, AG Reinhardt); der Temperatur-Gradient verlief von 50°C (auf Position 1) bis zu 70°C (auf Position 12).

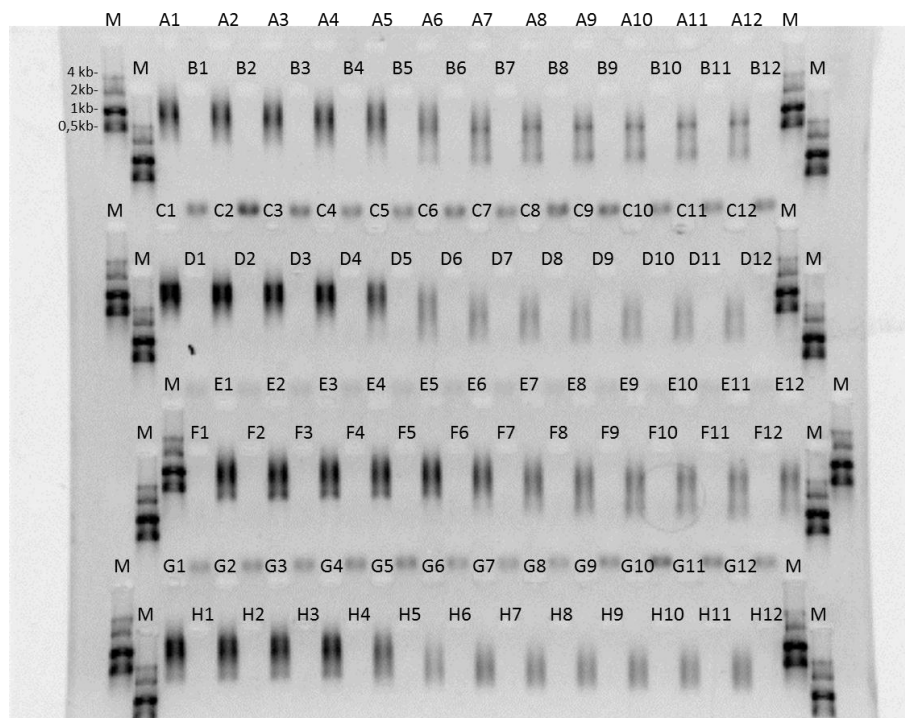


Abb. 51: Gelbild der Gradienten-PCR von aqf3b und aqf 4b

In Reihe A ist PCR-Ansatz 1 aufgetragen (aqf3b mit ADB-CDNA-Revers Primer), in Reihe C PCR-Ansatz 2 (aqf3b mit ADB-TANKER-Revers-Primer), in Reihe E PCR-Ansatz 3 (aqf4b mit ADB-CDNA-Revers-Primer), in Reihe G PCR-Ansatz 4 (aqf4b mit ADB-TANKER-Revers-Primer). Die Reihen B, D, F und H stellen die dazugehörigen Negativkontrollen dar.

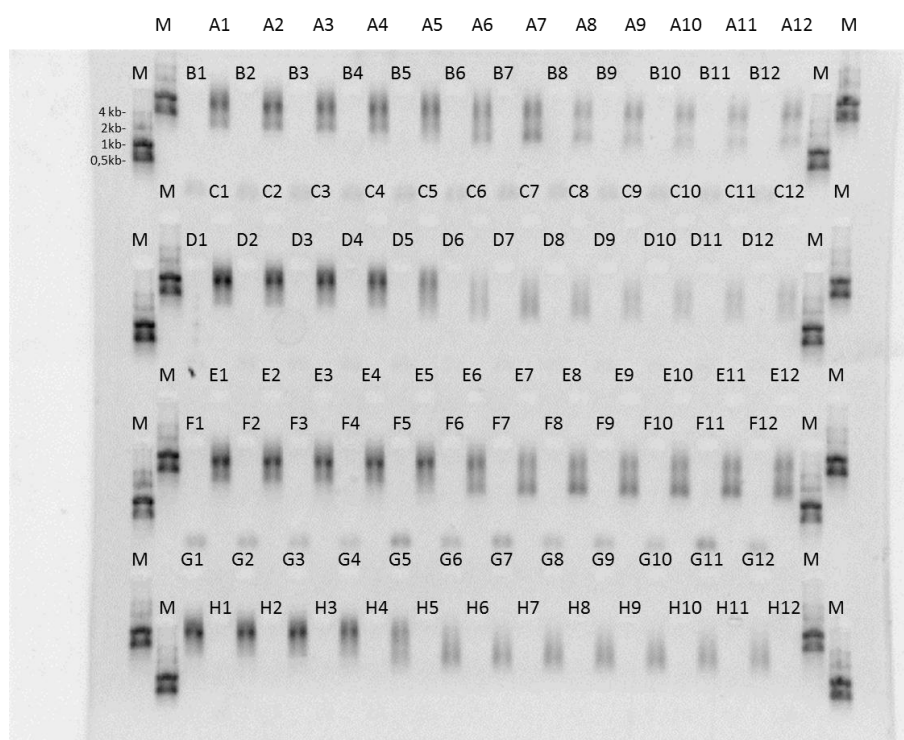


Abb. 52: Gelbild der Gradienten-PCR von aqf5b und aqf6b

In Reihe A ist PCR-Ansatz 5 aufgetragen (aqf5b mit ADB-CDNA-Revers Primer), in Reihe C PCR-Ansatz 6 (aqf5b mit ADB-TANKER-Revers-Primer), in Reihe E PCR-Ansatz 7 (aqf6b mit ADB-CDNA-Revers-Primer), in Reihe G PCR-Ansatz 8 (aqf6b mit ADB-TANKER-Revers-Primer). Die Reihen B, D, F und H stellen die dazugehörigen Negativkontrollen dar.

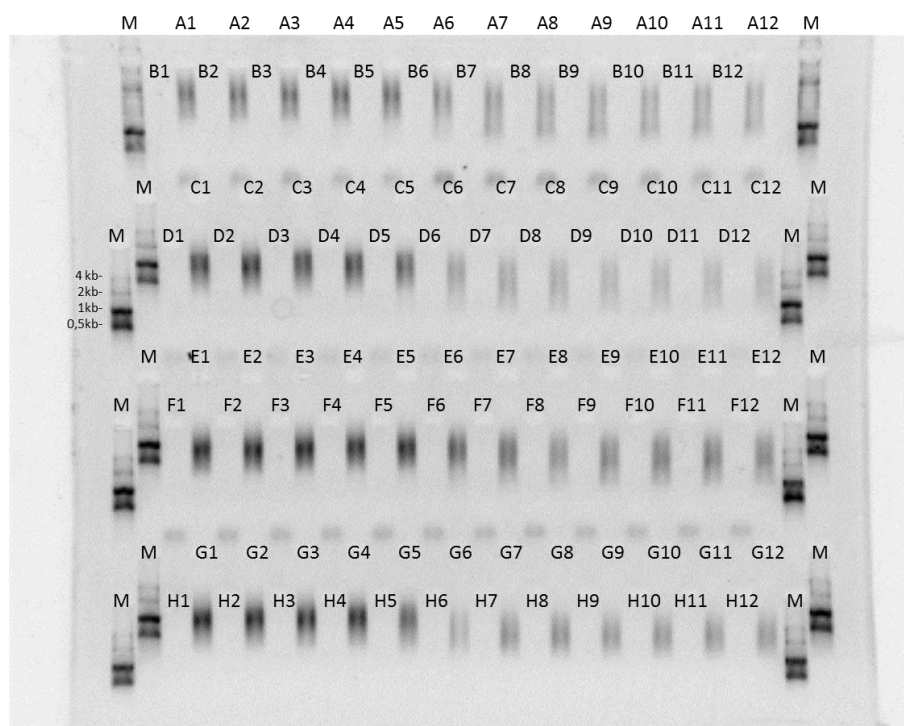


Abb. 53: Gelbild der Gradienten-PCR von aqf9b und aqf10b

In Reihe A ist PCR-Ansatz 9 aufgetragen (aqf9b mit ADB-CDNA-Revers Primer), in Reihe C PCR-Ansatz 10 (aqf9b mit ADB-TANKER-Revers-Primer), in Reihe E PCR-Ansatz 11 (aqf10b mit ADB-CDNA-Revers-Primer), in Reihe G PCR-Ansatz 12 (aqf10b mit ADB-TANKER-Revers-Primer). Die Reihen B, D, F und H stellen die dazugehörigen Negativkontrollen dar.

Auf allen Gelbildern sind PCR-Produkte der cDNAs in den Reihen zu sehen, es hat also eine Amplifikation der Template-DNA stattgefunden (die in der PCR eingesetzte cDNA-Menge von 1 ng pro 100 µl-Ansatz ist zu klein, als dass man sie auf den Ethidiumbromid-Gelen sehen könnte). Bei den entsprechenden Negativ-Kontrollen („B“- , „D“- , „F“- und „H“-Reihen) ist nur eine Primerbande erkennbar (außer bei den H-Reihe, dort sind die Banden aus dem Gel gelaufen), es fand also keine Amplifikation statt und die PCR-Ansätze waren frei von Kontaminationen. Die „A“- und „E“-Reihen stellen die Amplifikate der cDNAs mit dem ADB-CDNA-Revers-Primer dar, die „C“- und „G“-Reihen die mit dem ADB-TANKER-Revers Primer. Mit steigender Annealing-Temperatur (Position 1 = 50°C bis Position 12 = 70°C) werden die Amplifikationsprodukte aller cDNAs schwächer. Die Länge der Amplifikationsprodukte unterscheidet sich zwischen den beiden verwendeten Revers-Primern nicht bei entsprechenden Annealingtemperaturen (gleiche Spalte), allerdings scheinen die ADB-TANKER-Reversprimer-Produkte etwas schwächer zu sein, hier sind die Ausbeute etwas schlechter. Bei allen Ansätzen sind Produkte in einer Größe von 2 kb bis 0,5 kb vorhanden, das entspricht dem Längenbereich der ursprünglichen cDNAs; diese wurden also auch ihrer Länge entsprechend amplifiziert. Ab Position 6 (Annealing-Temperatur = 58°C) ändert sich bei allen Amplifikations-Produkten die Größenverteilung nicht mehr.

Für aqf3b wurde Position A6 (Abb. 51) der Gradienten-PCR mit Ampurebeads aufgereinigt und eine DNA-Konzentration von 22,75 ng/µl des Eluates bestimmt, für aqf4b auf Position E6 (Abb. 51) ergab sich nach Aufreinigung eine DNA-Konzentration von 14,03 ng/µl, für aqf5b (Abb. 52) auf Position A6 ergab sich nach Aufreinigung eine DNA-Konzentration von 10,71 ng/µl, für aqf6b auf Position E6 (Abb. 52) ergab sich nach Aufreinigung eine DNA-Konzentration von 10,28 ng/µl, für aqf9b auf Position A6 (Abb. 53) ergab sich nach Aufreinigung eine DNA-Konzentration von 19,92 ng/µl und für aqf10b auf Position E6 (Abb. 53) ergab sich nach Aufreinigung eine DNA-Konzentration von 28,84 ng/µl. Dies stellen die aqf-Amplicon Librarys dar, die bei einer geschätzten durchschnittlichen Länge von 1000 bp weiterverwendet werden (für Titration und dann Runstart).

3.2.2.2 Titration der Amplicon Librarys

Die sechs Amplicon Librarys wurden titriert, wobei jeweils cpb-Zahlen von 0,5, 1, 2 und 4 zur Anwendung kamen, um eine ideale cpb-Zahl für einen Run-Start zu bestimmen (Tab. 24 und Abb. 54 auf Seite 82).

Library	cpb	zurückgewonnene Beads	Beadrecovery- Rates in Prozent	aufgereinigte Beads	Enrichment-Rates in Prozent
aqf3b	0,5	794630	72,9	91740	11,5
	1	861040	79,0	155610	18,1
	2	874780	80,3	177156	20,3
	4	719060	66,0	225786	31,4
aqf4b	0,5	911420	83,6	38124	4,2
	1	712190	65,3	43938	6,2
	2	863300	79,2	56905	6,6
	4	881650	80,9	86020	9,8
aqf5b	0,5	615580	56,5	28837	4,7
	1	822110	75,4	32918	4,0
	2	716770	65,8	49875	7,0
	4	1174770	107,8	33718	2,9
aqf6b	0,5	657230	60,3	29019	4,4
	1	938900	86,1	27467	2,9
	2	785470	72,1	41200	5,2
	4	719060	66,0	71924	10,0
aqf9b	0,5	700000	64,2	79725	11,4
	1	735000	67,4	92250	12,6
	2	405000	37,2	82800	20,4
	4	775000	71,1	157250	20,3
aqf10b	0,5	705000	64,7	41920	5,9
	1	675000	61,9	68680	10,2
	2	695000	63,8	57715	8,3
	4	690000	63,3	117090	17,0

Tab. 24: Titration der Amplicon Librarys

Die Tabelle zeigt die mittels Partikelzählers gewonnenen Messergebnisse der Beads nach Beadrecovery (zurückgewonnene Beads) und Enrichment (aufgereinigte Beads) und die sich daraus ergebenden prozentualen Recovery-Rates. Für jede cpb-Zahl wurde bei der Beadmessung eine Vierfachbestimmung durchgeführt; in der Tabelle sind die entsprechenden Mittelwerte angegeben.

Die Beadrecovery-Rates liegen mit wenigen Ausnahmen zwischen 65% und 80%, also in dem von Roche als typisch bezeichneten Bereich für das Recovery über Spritzen [122]. Bei den meisten Librarys steigen die Enrichment-Rates mit steigenden cpb-Werten, was ebenfalls zu erwarten war.

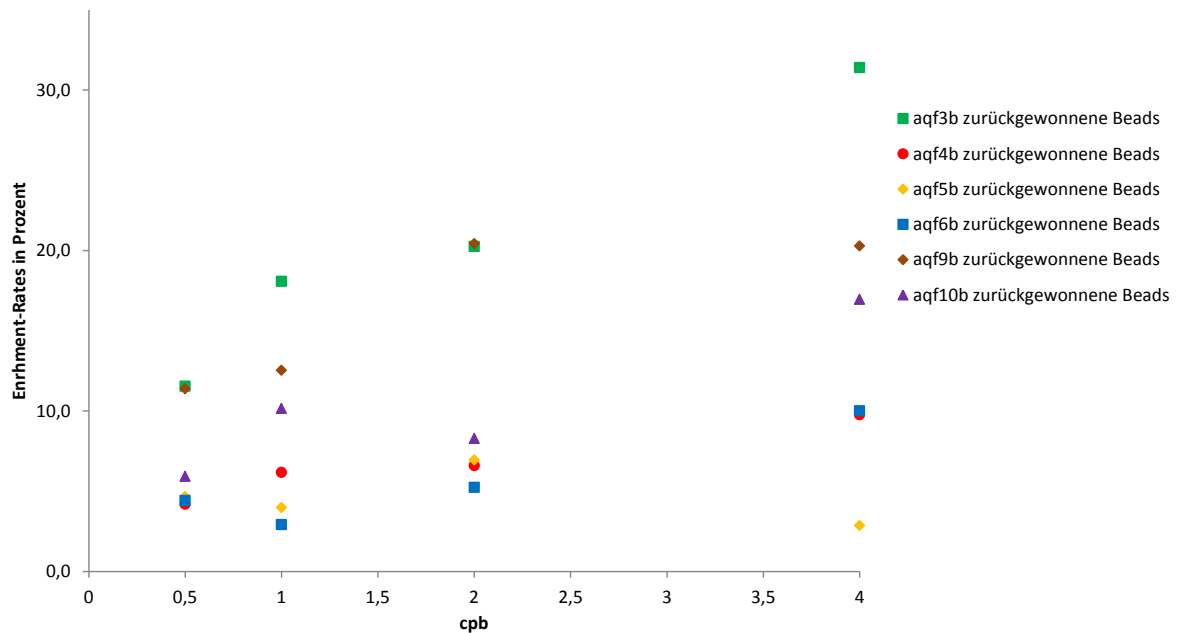


Abb. 54: Graphische Darstellung der Amplicon Library-Titrationsergebnisse

Aqf5b zeigt einen atypischen Verlauf (Abb. 54); bei 4 cpb weist die Library einen Enrichment-Wert von nur 3% auf, besitzt aber eine Beadrecovery-Rate von über 100%. Diese kann entweder aufgrund von Pipettierfehlern bei der Capture-Bead-zugabe zustande kommen, oder aber aufgrund von Fehlern bei der Messung mit dem Partikelzähler (zu viele Beads in die Messung pipettiert, oder zu starkes Schütteln, das Luftblasenbildung verursacht). Nimmt man letzteres an, so ist die Beadrecovery-Rate vermutlich kleiner, was wiederum auch eine größere Enrichment-Rate bedeutet, und der Messkurven-Verlauf bei 4 cpb nicht mehr so stark abflachen würde. Über die Bestimmung der Enrichment-Rate soll für jede Library ein geeigneter cpb-Wert gefunden werden, der eine ausreichende Menge an Beads bedeutet, aber auch unterhalb der von Roche angegebenen Obergrenze von 15-20% liegt. Anhand der Messergebnisse wurde für aqf3b und aqf9b ein cpb-Wert von 0,5 festgelegt, für aqf4b und aqf6b ein cpb-Wert von 4 und für aqf10b ein cpb-Wert von 2. Für aqf5b wurde ein cpb-Wert von 6 festgelegt, da die Enrichment-Rate von 4 cpb aufgrund der obigen Überlegungen als zu niedrig angenommen wurde; bei einer angestrebten Enrichment-Rate von um die 10% sollte man mit 6 cpb in die Nähe kommen, wenn man in dem Bereich einen linearen Anstieg annimmt.

3.2.2.3 Beadrecovery der Amplicon Librarys

Mit den sechs Amplicon Librarys wurde anhand der aus den Titrationsergebnissen erhaltenen cpb-Werten eine große emPCR angesetzt und ein Beadrecovery durchgeführt (Tab. 25), wobei jede Library getrennt von den anderen angesetzt und amplifiziert wurde (siehe Tab. 11 auf Seite 51).

	Librarys	zurückgewonnene Beads	Beadrecovery- Rates in Prozent	aufgereinigte Beads	Enrichment-Rates in Prozent
Cup A	aqf3b	9362050	80,2	531200	5,7
Cup B	aqf4b	10121030	86,8	1469120	14,5
Cup C	aqf5b	9590987	82,2	985255	10,3
Cup D	aqf6b	9250683	79,3	805120	8,7
Cup E	aqf9b	9807852	84,1	1650500	16,8
Cup F	aqf10b	9690652	83,1	845000	8,7

Tab. 25: Beadrecovery der Amplicon Librarys

Es werden die Messergebnisse der Beads nach Beadrecovery (zurückgewonnene Beads) und Enrichment (aufgereinigte Beads) gezeigt und die daraus berechneten prozentualen Recovery-Rates.

Bei fast allen Librarys wurden die Beads mit Beadrecovery-Rates von über 80% aus den Emulsionen zurückgewonnen, wobei diese nahe der von Roche angegebenen Rate von 85% liegen. Die Enrichment-Rates liegen bei fünf Librarys unterhalb des Grenzbereiches von 15-20%, nur aqf9b liegt mit knapp 17% innerhalb des Grenzbereiches. Die Enrichment-Rates aller Librarys/Cups liegen nahe an den bei der Titration angepeilten Werten von um die 10%. Bis auf aqf3b wurden bei allen anderen Librarys die Marke von 700000 Beads an aufgereinigten Beads überschritten –bei sechs Librarys, die zu gleichen Teilen in einem Run sequenziert werden sollen, sind rund 700000 Beads pro Library nötig.

3.2.2.4 Sequencing Run der Amplicon Librarys

Von den sechs aqf-Amplicon Librarys wurde ein große emPCR angesetzt, diese im Beadrecovery aufgereinigt und mit den Beads ein Sequencing-Lauf durchgeführt. Tab. 26 (Seite 84) zeigt die Ergebnisse des Runs, in Abb. 55 bis Abb. 57 sind sie graphisch dargestellt; in Tab. 27 und Abb. 58 sind die Ergebnisse der Kontroll-DNA-Analyse aufgeführt.

Bei diesem Sequenzierlauf wurden nur insgesamt 163305 Reads (Sequenzen mit TCAG-Key) produziert (Passed Filter Wells, Total), was einem Anteil von 11% (Passed Filter Wells) an den Key-Pass-Wells entspricht (Tab. 26). Mit rund 750000 erkannten Raw Wells auf Region 1 und 800000 Raw Wells auf Region 2 liegt dieser Run etwas unterhalb der Roche-Vorgaben (bei einem guten Run werden 1-1,2 Mio. Raw Wells pro Region erwartet). Die durchschnittliche Länge dieser Sequenzen ist mit 150 bp sehr kurz (Lenght Average Total), wobei der Großteil der Reads zwischen 75 und 175 bp liegt, es gibt kaum eine Häufung von Reads größer als 300 bp (Abb. 55). Beides liegt deutlich unterhalb der Vorgaben von Roche; für einen Ampliconlauf werden ca. 375000 Reads pro Region erwartet [94] mit Längen zwischen 300-400 bp (von der Amplicongröße abhängig). Auch dieser Sequenzier-Lauf ist sehr schlecht, er zeigt sogar noch geringere Ausbeuten als der Lauf mit den aqf-General-Librarys.

Der Anteil an Reads, der herausgefiltert wurde, ist sehr hoch; rund 43% blieben im Dot & Mixed Filter hängen und 45% waren zu kurz (% short, Total, Tab. 26). Den größten Anteil an den Failed Reads haben die, die als „too Short“ aussortiert wurden, dieser liegt mit 559324 Reads ungewöhnlich hoch („Failed Short Quality“, Total, siehe Tab. 26 und Abb. 57); gefolgt von Reads, die vom Dot-Filter entfernt wurden –darauf entfallen 429972 Reads insgesamt. 106634 Reads wurden bei dem Lauf als „Failed Short Primer“ herausortiert (Tab. 26), was sowohl von ihrer Anzahl als auch von dem Anteil

an den Reads hoch ist (Abb. 57) –bei guten Sequenzierläufen ist die Zahl an diesen Reads verschwindend gering bezogen auf die Gesamtzahl an Reads und liegt unter 2000 Reads insgesamt pro Sequenzier-Lauf.

Region 1 ist in diesem Run noch etwas schlechter als Region 2, es kamen mit 65868 Reads weniger herunter als auf Region 2 (97437 Reads), sie sind im Durchschnitt mit 129 bp noch kürzer als die von Region 2 (165 bp) und dementsprechend war der Anteil an Failed Reads auf Region 1 etwas größer als bei Region 2, wobei auf Region 1 vorallendingen noch mehr Reads als „Failed Short Quality“ herausortiert wurden (303773 Reads gegenüber 255551 Reads auf Region 2).

Die Kontroll-DNA hat auf Region 1 rund 83% Passed Filter Wells und auf Region 2 82% (CATG-Wells, Tab. 26), was innerhalb der Vorgaben für einen guten Lauf liegt. Mit durchschnittlichen Leselängen von rund 400 bp auf Region 1 und 380 bp auf Region 2 und einer entsprechenden Größenverteilung (Abb. 56) sind die Reads der Kontroll-DNA unauffällig und entsprechen ebenfalls den Vorgaben für einen guten Lauf. Genauso sieht es mit der Sequenziergenauigkeit der Kontroll-DNA-Reads aus (Tab. 27 und Abb. 58); eine Übereinstimmung von 95% über eine Länge von 100 bp erreichen auf beiden Regionen mehr als 82% der Kontroll-DNA-Reads; eine Übereinstimmung von 95% über 200 bp Länge werden von rund 77% erreicht, über 300 bp von 75% und über 400 bp von immer noch rund 62%. Diese Werte sind zwar nicht sehr gut, aber sind auch noch nicht auffällig (mündl. Rücksprache mit dem Roche-Support).

	TCAG-Wells (Library)			CATG-Wells (Control)		
	Region 1	Region 2	Total	Region 1	Region 2	Total
Raw Wells	745578	801667	1547245	745578	801667	1547245
Key Pass Wells	719440	753997	1473437	3171	5076	8247
Passed Filter Wells	65868	97437	163305	2616	4111	6727
Total Bases	8501833	16113810	24615643	1043915	1563478	2607393
% Passed Filter	9,16	12,92	11,08	82,50	80,99	81,57
Lenght Average	129,07	165,38	150,73	399,05	380,32	387,6
Lenght Std Deviation	81,78	111,32		131,07	138,05	
Longest Reads Lenght	709	931	931	564	619	619
Shortest Reads Lenght	40	40	40	40	40	40
Median Reads Lenght	103	124	118	446	407	432
Failed Dot	202037	227935	429972	21	19	40
Failed Mixed	93732	107950	201682	16	54	70
Failed Short Quality	303773	255551	559324	485	842	1327
Failed Short Primer	48459	58175	106634	0	0	0
% Dot + Mixed	41,11	44,55	42,87	1,17	1,44	1,33
% Short	48,96	41,61	45,20	15,29	16,59	16,09

Tab. 26: Run mit den aqf-Amplicon Librarys

Es sind die Ergebnisse des Runs dargestellt, die von der GSRUNBrowser-Software ausgegeben werden; links in der Tabelle befinden sich die der Library-DNA (tcag-wells) und rechts die der beim Sequencing mitgeladenen Kontroll-DNA (catg-wells).

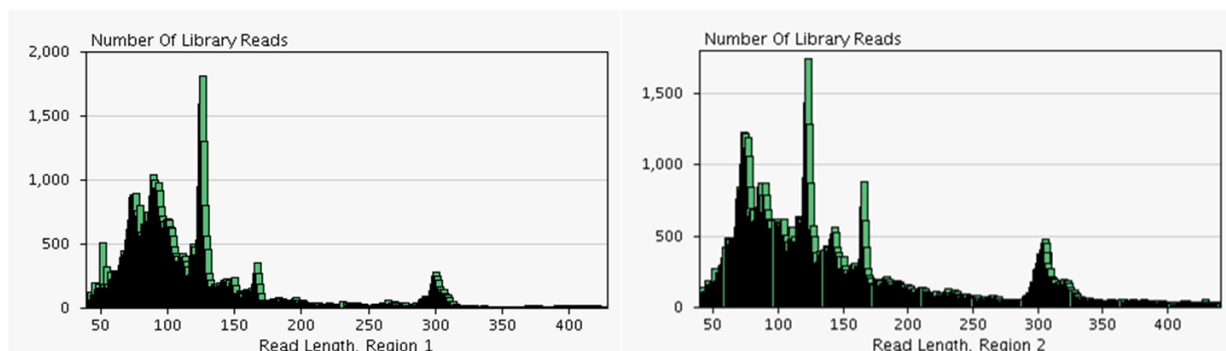


Abb. 55: Größenverteilung der Reads beim Run mit den aqf-Amplicon Libraries

Es werden die von der GSRUNBrowser-Software ausgegebenen Verteilungen der Readlängen von Sequenzen mit tcag-Key auf Region 1 (links) und Region 2 (rechts) gezeigt.

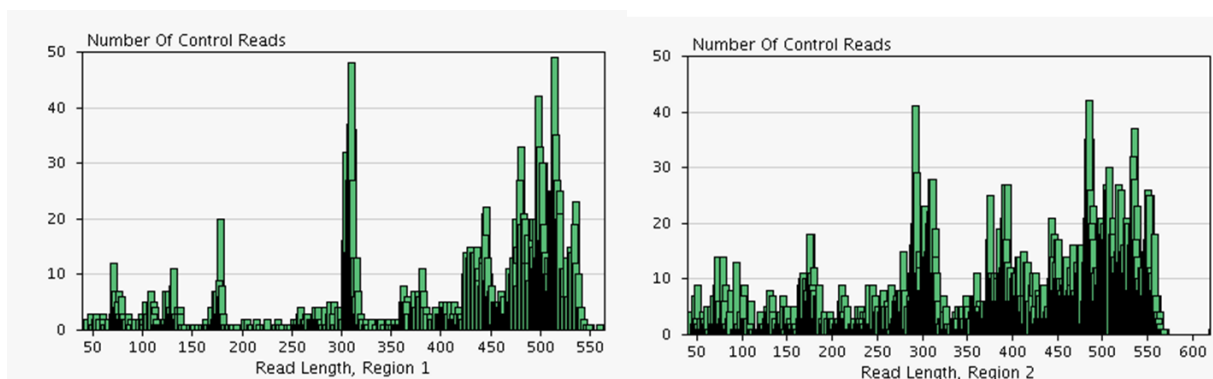


Abb. 56: Größenverteilung der Kontroll-DNA beim Run mit den aqf-Amplicon Libraries

Es werden die von der GSRUNBrowser-Software ausgegebenen Verteilungen der Kontroll-DNA-Readlängen (Reads mit dem catg-Key) auf Region 1 (links) und Region 2 (rechts) gezeigt.

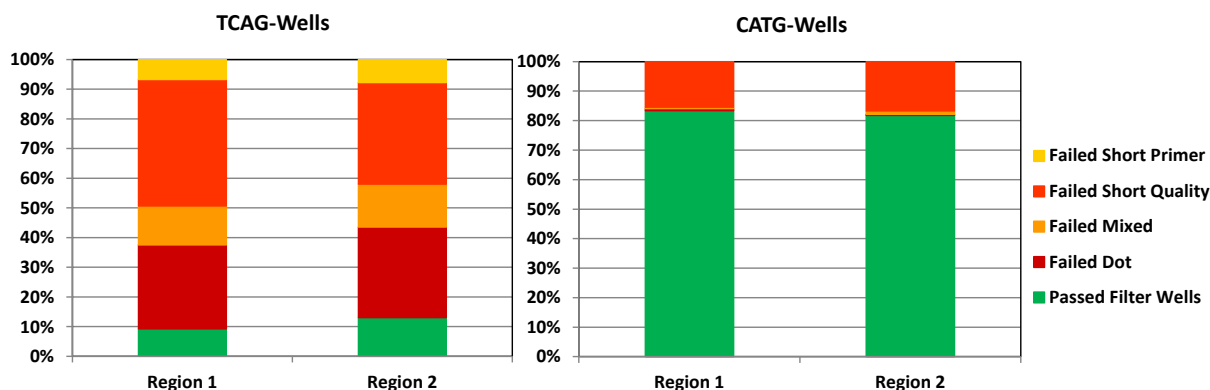


Abb. 57: Prozentuale Verteilung beim Run mit den aqf-Amplicon Libraries

Es wird das Verhältnis auf beiden Regionen von Passed Filter Wells (grün) zu den Failed Wells dargestellt, angezeigt für Library-Reads (tcag-wells) und Kontroll-DNA-Reads (catg-wells).

Länge	100 bp			200 bp			300 bp			400 bp		
	100%	98%	95%	100%	98%	95%	100%	98%	95%	100%	98%	95%
Region 1	50,58	74,71	83,63	41,19	70,36	77,17	25,58	67,87	75,50	12,68	56,10	62,60
Region 2	47,89	74,53	82,35	39,46	69,58	76,97	21,99	66,96	75,10	10,76	54,94	61,86

Tab. 27: Kontroll-DNA-Analyse beim Run mit den aqf-Amplicon Libraries

Es sind die von der GSRUNBrowser-Software ausgegebenen Ergebnisse der Kontroll-DNA-Analyse dargestellt. Für beide Regionen des Runs wird der prozentuale Anteil der Kontroll-Bead-Reads angegeben, der zu je 100%, 98% und 95% mit den Kontroll-DNA-Sequenzen über eine Länge von 100 bp, 200 bp, 300 bp und 400 bp übereinstimmt (Match).

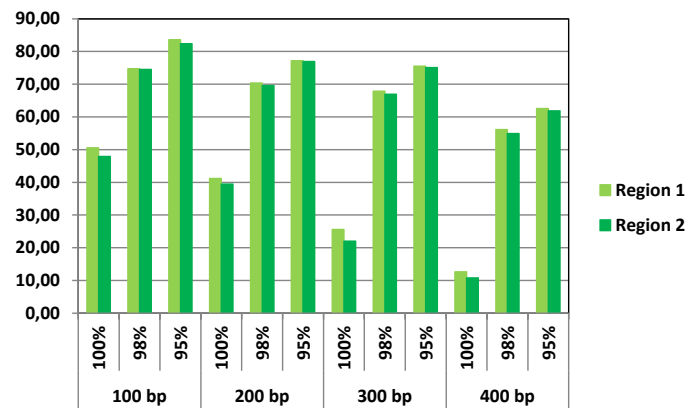


Abb. 58: Graphische Darstellung Kontroll-DNA-Analyse

Es werden die Ergebnisse der Kontroll-DNA-Analyse aus Tab. 27 graphisch dargestellt.

Die Sequenzierung der aqf-Amplicon Librarys schlug fehl, die Run-Ausbeuten sind katastrophal gering (Passed Filter Wells) und die durchschnittlichen Leseweiten viel zu klein. Diese Amplicon Librarys können nicht weiter verwendet werden. Zur Ursachenforschung wurden Amplifikations-Tests mit den Amplicon Librarys durchgeführt; die Ergebnisse werden in den folgenden Kapiteln dargestellt.

3.2.2.5 Amplifikations-Test mit den aqf-Amplicon Librarys

Für jede der sechs Amplicon Librarys wurde ein Amplifikations-Test angesetzt, wobei als Primer die Adapter-A- und -B-Sequenzen verwendet wurden (vergleiche Seite 52); Aliquots der PCR wurden auf ein Gel aufgetragen (Abb. 59). Damit sollte es möglich sein einen ersten Eindruck zu erhalten, ob die aqf-Amplicon Librarys überhaupt „echte“ 454-Librarys darstellen.

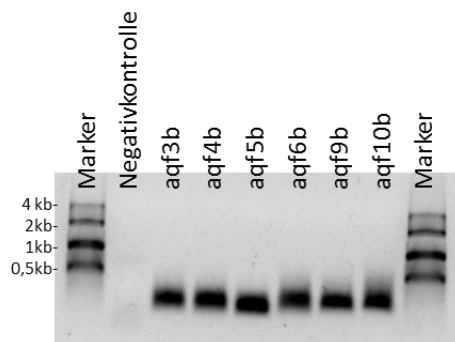


Abb. 59: Gelbild des Amplifikations-Tests

Der Marker mit seinen 4kb-, 2kb-, 1kb- und 0,5kb-Banden ist eine Eigenproduktion des MPI für molekulare Genetik, AG Reinhardt.

Bei der Negativkontrolle ist den Schatten der Primerbande zu erkennen, außerdem vielleicht eine sehr schwache Bande kurz oberhalb von 0,5 kb. Alle sechs aqf-Amplicon Librarys zeigen sehr kurze Amplifikationsprodukte mit teilweise leichten DNA-Schmier hoch zu 0,5 kb.

Auf der einen Seite gibt es tatsächlich echte Amplifikationsprodukte, deren Bandenlängen die Ergebnisse des Sequencing-Runs widerspiegeln, auf der anderen Seite stehen die Bandenlängen im

Widerspruch zu den Ergebnissen der Gelbilder der aqf-Amplicon Librarys (Abb. 51 bis Abb. 53 auf den Seiten 78 und 79) –die dortigen Banden stellen zwar Amplifikations-Produkte dar, allerdings scheinen sie keine fehlerfreien Adapter-A- und –B-Bereiche zu besitzen, sonst hätten die Banden dieses Amplifikationstests größer sein müssen. In weiteren folgenden Tests sollte versucht werden die Ursache hierfür zu ergründen, welche im folgenden Abschnitt beschrieben sind.

3.2.2.6 Amplifikations-Tests mit aqf-Sanger-Klonen

Um den Multiplex-Charakter der aqf-cDNAs als Fehlerquelle auszuschließen, wurden aus aqf-cDNA-Sanger-Klonen 454-Amplicon Librarys erstellt.

Mit aqf-cDNA-Sangerklonen wurde eine Gradienten-PCR mit den ADA-MID1-CDNA- und ADB-CDNA-Primern durchgeführt (vergleiche Seite 52 f) und die Amplifikationsprodukte auf ein Agarosegel aufgetragen (Abb. 60).

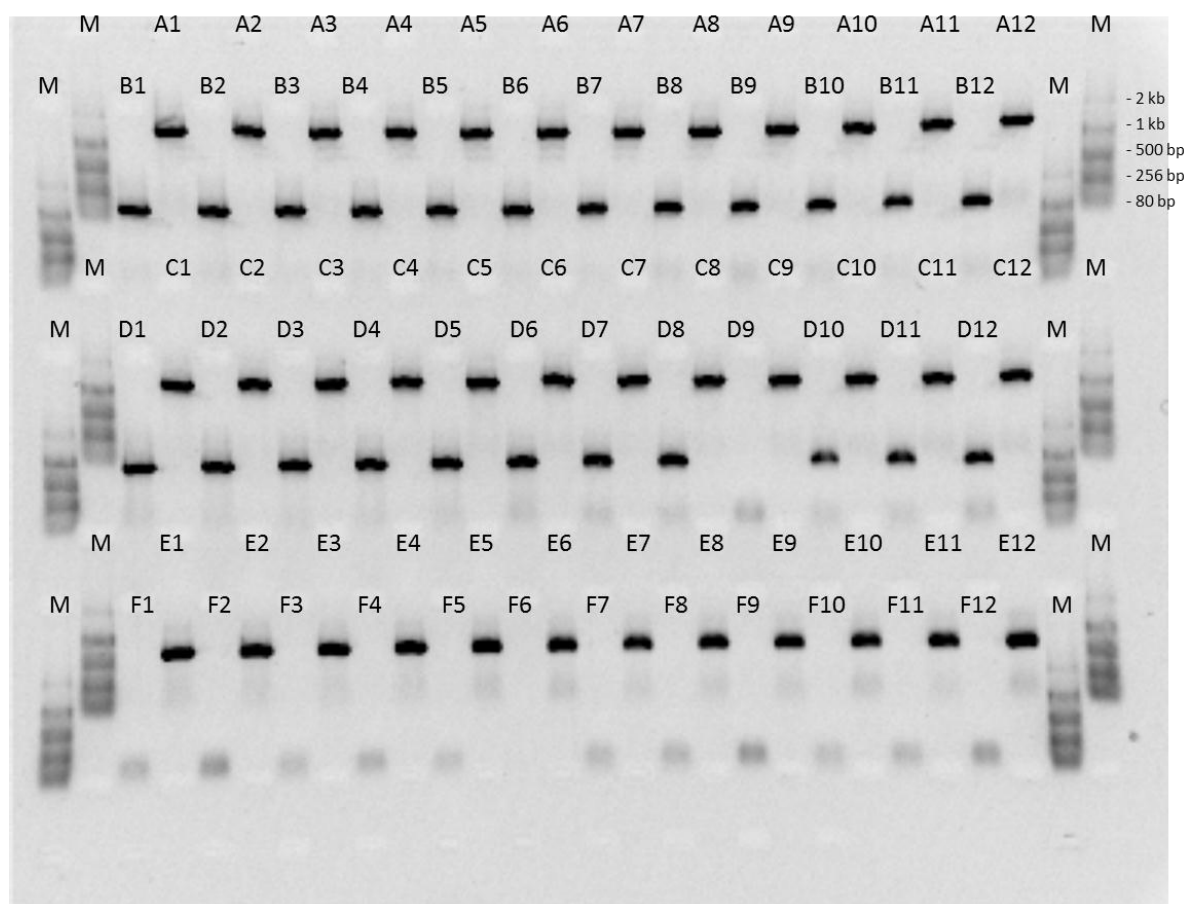


Abb. 60: Erstellung von Amplicon Librarys mit aqf-Sanger-Klonen

In Reihe „A“ stellte der Sangerklon aqf3b_1 die Template-DNA dar, in Reihe „B“ aqf3b_3, in Reihe „C“ aqf3b_4, in Reihe „D“ aqf3b_4 und in Reihe „E“ aqf3b_8. Reihe „F“ ist die Negativ-Kontrolle, der Temperatur-Gradient beim Annealing in der PCR verlief von 50°C (Spalte „1“) bis zu 70°C (Spalte „12“). Mit „M“ sind die Marker-Lanes gekennzeichnet; der Marker ist eine Eigenproduktion des MPI für molekulare Genetik und besitzt Banden bei 2 kb, 1 kb, 500 bp, 256 bp und 80 bp.

Bei allen Sanger-Klonen (Reihen A bis E) sind Banden in einer Größe um 1 kb vorhanden (Ausnahme ist die Position D9). Es hat also eine Amplifikation von Template-DNA stattgefunden, da die Template-Konzentration in den 100 µl-Ansätzen zu klein ist (20 ng pro 100 µl), um sie auf einem Ethidium-

Bromid-Gel in den aufgetragenen 5 µl (entspricht 1 ng) zu sehen. Die Markerbanden auf dem Gel besitzen ca. 50 ng pro Bande, weshalb die Menge der Amplifikationsprodukte größer ist, da diese stärkere Banden zeigen. In den Negativ-Kontrollen (Reihe „F“) sind sehr dünne Banden bzw. Bandenschatten bei ca. 80 bp zu sehen.

Um festzustellen, ob die Amplicon Librarys der Sangerklone die Adapter-A- und -B-Bereiche tragen, wurde mit den Amplifikationsprodukten der Gradienten-PCR eine weitere PCR durchgeführt, wobei die 454-Adapter-A- und -B-Sequenzen als Primer zum Einsatz kamen; 5 µl Aliquots der PCR-Produkte wurden auf ein Agarosegel aufgetragen (Abb. 61).

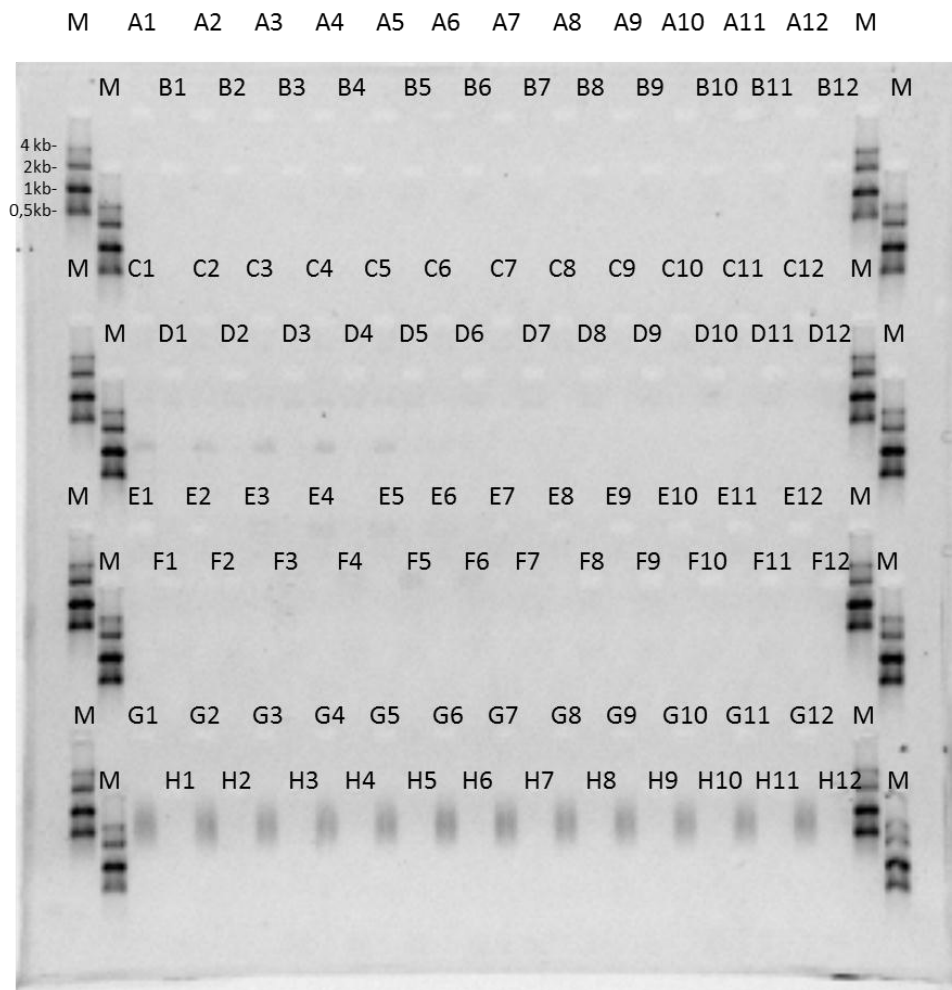


Abb. 61: Amplifikation der Gradienten-PCR-Produkte mit 454-Primern

Die Reihen „A“ bis „F“ stellen die Amplifikations-Produkte der Gradient-PCR der dortigen Positionen dar; Reihe „G“ ist die Positiv-Kontrolle und Reihe „H“ die Negative-Kontrolle. Mit „M“ sind die Marker-Lanes gekennzeichnet; der Marker ist eine Eigenproduktion des MPI für molekulare Genetik und besitzt Banden bei 4 kb, 2 kb, 1 kb und 500 bp.

Auf dem Gelbild (Abb. 61) sind in den Reihen „A“ bis „F“ keine Amplifikations-Produkte mit einer Länge um die 1 kb zu erkennen. Bei den Ansätzen C1 bis C5 sind schwache Banden von weit unter 500 bp zu sehen, alle anderen Ansätze der Amplicon Librarys aus den aqf-cDNA-Sanger-Klonen enthalten kein Amplifikationsprodukt. Die Positiv-Kontrollen in Reihe „G“ zeigen einen DNA-Schmier im erwarteten Größenbereich von 500 bp; die Negativkontrollen (Reihe „H“) zeigen kein Produkt. Auch

wenn in fast allen Reihen keine Amplifikations-Produkte zu sehen sind, hat die PCR funktioniert, da die Negativkontrollen frei von Produkten sind und die Positiv-Kontrolle Produkte zeigen, die dem Größenbereich der 454-Library-DNA entsprechen. Bei den Sangerklonen zeigt sich also ein entsprechendes Bild; die Amplicon Library besitzt keine funktionsfähigen Adapter-A- und -B-Bereiche. Der Multiplex-Charakter der aqf-cDNA kann damit als Fehlerquelle ausgeschlossen werden, da auch mit monoklonalen Sangerklonen keine funktionierende 454-Library erstellt werden kann.

Eine andere mögliche Fehlerquelle könnten die Primer-Overhangs darstellen, deren lange 5'-Enden auch noch „innerhalb“ der Template-DNA als Overhang vorliegt (vergleiche Abb. 36 auf Seite 49). Als Alternative sollte in einer Two-Step-PCR versucht werden, in einem ersten Schritt ein besser geeignetes Template zur Verfügung zu stellen, und in einem zweiten Schritt die benötigten 454-Adapter-Bereiche heranamplifiziert werden (Kapitel 2.2.6.10 auf Seite 52). Dazu wurden Sangerklone in einer ersten PCR unter Verwendung von Shortfor- und Shortrev-Primer amplifiziert (Abb. 62) und in einer zweiten PCR fand eine Verlängerung dieser Produkte statt (Abb. 63).

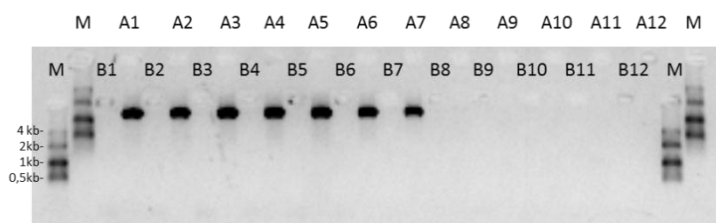


Abb. 62: Two-Step-PCR -erste Amplifikation

Reihe „A“ zeigt die Amplifikationsprodukte von Sangerklon aqf3b_1, Reihe „B“ die entsprechende Negativkontrolle. Der Temperatur-Gradient beim Annealing in der PCR verlief von 50°C (Spalte „1“) bis zu 70°C (Spalte „12“). Mit „M“ sind die Marker-Lanes gekennzeichnet; der Marker ist eine Eigenproduktion des MPI für molekulare Genetik und besitzt Banden bei 4 kb, 2 kb, 1 kb und 500 bp.

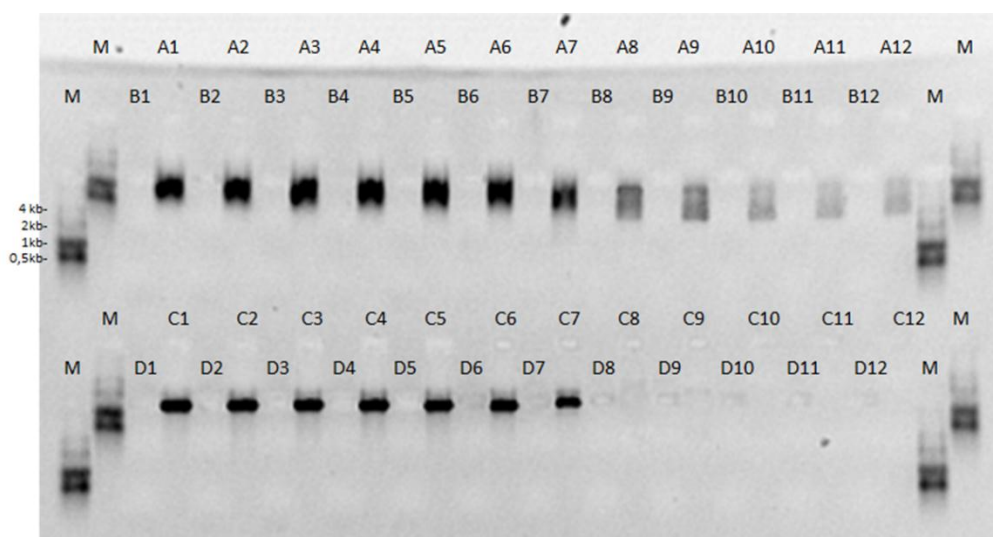


Abb. 63: Two-Step-PCR –zweite Amplifikation

Reihe „A“ zeigt die Positivkontrolle; Reihe „B“ die Negativkontrolle und Reihe „D“ die Amplifikationsprodukte mit der Template-DNA aus der ersten Amplifikation der Two-Step-PCR. In Reihe „C“ sind als Vergleich 5 µl der Template-DNA der ersten Amplifikation aufgetragen (Reihe „A“ aus Abb. 62). Der Temperatur-Gradient beim Annealing in der PCR verlief von 50°C (Spalte „1“) bis zu 70°C (Spalte „12“). Mit „M“ sind die Marker-Lanes gekennzeichnet; der Marker ist eine Eigenproduktion des MPI für molekulare Genetik.

Der erste Amplifikationsschritt (Abb. 62) der Two-Step-PCR zeigt, dass eine Amplifikation des Sangerklons (Reihe „A“) in der Gradienten-PCR erfolgreich war, es ist ein Amplifikat in der richtigen Größe (kurz oberhalb von 1 kb) zu sehen, das außerdem eine Bande ist. Bei höheren Annealingtemperaturen (ab Position A8) ist kein Produkt mehr zu sehen (entspricht 62°C) –die verwendeten kurzen Primer konnten dort wohl nicht mehr an die Template-DNA binden. Die Negativkontrolle (Reihe „B“) ist frei von Produkt.

Auf die Produkte des ersten Amplifikationsschrittes wurde eine erneute PCR aufgesetzt; Reihe „D“ des Gelbildes (Abb. 63) zeigt die Ergebnisse: es fand keine Amplifikation statt, es ist kein PCR-Produkt zu sehen; in Reihe „C“ sind als Vergleich nochmal die Amplifikationsprodukte des ersten Two-Step-Schrittes zu sehen. Die Negativkontrolle (Reihe „B“) ist frei von Produkt und die Positivkontrolle in Reihe „A“ hat funktioniert, dort sind PCR-Produkte in dem erwarteten Größenbereich zu sehen; das PCR-Mix scheint also in Ordnung gewesen zu sein.

Auch das verwendete Two-Step-PCR-Protokoll führte nicht zum Erfolg. Das Template nach der ersten PCR, das theoretisch besser für die Verlängerung um die 454-Adapter-Sequenzen geeignet sein sollte, zeigt kein Amplifikationsprodukt. Als mögliche Fehlerquelle bei der Amplicon Library-Erstellung aus den aqf-cDNAs bleiben mögliche Primer-Inkompatibilitäten zwischen 454-Sequenzen auf der einen Seite und Template-DNA auf der anderen Seite (MINT-Sequenzbereiche oder die cDNA selber). Da aber weder die 454-Sequenzen, noch die Template-DNA geändert werden kann, musste eine neue Methode zur 454-Library-Erstellung entwickelt werden, was in den folgenden Abschnitten beschrieben ist.

3.3 Entwicklung einer neuen Methode zur gerichteten Sequenzierung

Es wurde versucht, eine neue Methode zur Library-Herstellung zu entwickeln, die das gerichtete Sequenzieren von Insert-DNA ermöglicht; das Prinzip ist in Abb. 37 auf Seite 54 dargestellt.

3.3.1 Erprobung einer Testlibrary

In ersten Versuchen sollte mit der neuen Methode eine Library erstellt, diese geprüft und dann sequenziert werden. Um die anfallenden Kosten des 454-Sequenzierens bei einem möglichen Ausfall der so erstellten Library gering zu halten, sollte diese in einem Run mit anderen Proben, die nach dem General-Library-Protokoll hergestellt wurden, in einem kleinen Anteil untergemischt werden (vergleiche Seite 54 f.).

3.3.1.1 Erstellung der Testlibrary

Nach der doppelten SPRI-Aufreinigung wurde die Konzentration der aqf65-Testlibrary mittels Fluorometer (Qubit) bestimmt. Diese lag bei 62,3 ng/μl. Mit 0,5 μl davon wurde eine Test-PCR mit den Adapter-A- und -B-Sequenzen als Primer durchgeführt (Abb. 64).

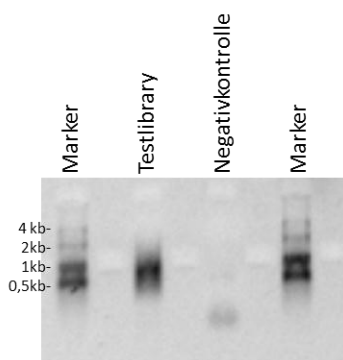


Abb. 64: Gelbild der PCR mit der Testlibrary

Das Gelbild zeigt die Test-PCR mit den Adapter-A- und -B-Sequenzen als Primern von Testlibrary und Negativkontrolle. Der Marker ist eine Eigenproduktion des MPI für molekulare Genetik und besitzt Banden bei 4 kb, 2 kb, 1 kb und 500 bp.

Die Test-Library zeigt Amplifikationsprodukte als Schmier im Bereich von 0,5 kb bis hoch zu ca. 1,5 kb. Damit liegen die Amplifikationsprodukte in dem erwarteten Größenbereich der aqf-cDNAs. Die Negativkontrolle zeigt eine leichte Bande bei weit unterhalb von 500 bp; dabei kann es sich um Primer oder möglicherweise sehr kurze Produkte handeln.

Anhand der Größenverteilung der Amplifikationsprodukte der Test-Library wird die durchschnittliche Größe auf 1 kb geschätzt.

3.3.1.2 Beadrecovery der emPCR mit der Test-Library

Die Test-Library wurde in einer großen emPCR untergemischt; die unterschiedlichen Librarys wurden dabei also nicht getrennt angesetzt. Nach Amplifikation in der emPCR fand ein Beadrecovery statt (Tab. 28).

Librarys	zurückgewonnene Beads	Beadrecovery- Rates in Prozent	aufgereinigte Beads	Enrichment-Rates in Prozent
271-15				
mge011				
Test-Library	29405602	84,0	2600020	8,8
FR7				
FR8				

Tab. 28: Beadrecovery der emPCR mit der Test-Library

Es werden die Messergebnisse der Beads nach Beadrecovery (zurückgewonnene Beads) und Enrichment (aufgereinigte Beads) gezeigt und die daraus berechneten prozentualen Recovery-Rates. Es ist nur der Cup, in dem die Test-Library untergemischt wurde, angegeben.

Die Beads wurden mit einer Ausbeute von 84% aus der Emulsion zurückgewonnen, womit man den Roche-Vorgaben von erwarteten 85% entspricht. Die Enrichment-Rate ist mit 8,8% ebenfalls gut; sie liegt deutlich von dem Grenzbereich von 15-20% entfernt, wobei aber genug Beads für einen Laufstart für eine Region erzeugt wurden (über 2 Mio.).

3.3.1.3 Sequencing-Run mit der Test-Library

2,2 Mio. aufgereinigter Beads, die die Test-Library enthalten sollen, wurden auf Region 1 der PTP geladen und sequenziert (die Ergebnisse der anderen Region werden nicht dargestellt). In Tab. 29 und den Abbildungen auf den folgenden Seiten sind die Ergebnisse des Runs dargestellt; Tab. 30 und Abb. 67 zeigen eine Analyse der Kontroll-DNA. Tab. 31 beschreibt die Zerlegung der erzeugten Reads nach ihren MIDs, Abb. 68 zeigt die Größenverteilung der Test-Library.

Auf Region 1 wurden über 550000 Reads (Sequenzen mit TCAG-Key) produziert (Passed Filter Wells) mit einem Anteil von knapp über 60% (Passed Filter Wells) an den Key-Pass-Wells (Tab. 29). Die durchschnittliche Länge der Reads liegt bei 365 bp, was sich in einer entsprechenden Größenverteilung äußert (Abb. 65). Damit liegt Region 1 innerhalb der Vorgaben von Roche für einen guten Lauf (über 400000 Reads pro Region mit einer durchschnittlichen Leseweite von über 350 bp). Auch die Raw Wells entsprechen mit knapp einer Mio. den Vorgaben. Dementsprechend ist der Anteil an herausgefilterten Reads klein, wobei der Großteil (249176 Reads) als „Failed Short Quality“ aussortiert wurde. Mit insgesamt 12,73% blieben nur wenig Reads im Dot & Mixed Filter hängen und als „Failed Short Primer“ wurden lediglich 279 Reads heraussortiert (vergleiche Tab. 29 bzw. Abb. 66). Die Kontroll-DNA-Reads zeigen ebenfalls gute Ergebnisse: die durchschnittliche Leselänge beträgt 385 bp bei einer entsprechenden Größenverteilung (Abb. 65); die Ausbeuten sind mit knapp 88% Passed Filter Wells hoch (Abb. 67) und die prozentualen Anteile mit 100%iger, 98%iger und 95%iger Übereinstimmung über alle Längen gut (Tab. 30 und Abb. 67).

Eine Zerlegung der Reads nach ihren MIDs ergibt, dass die Test-Library mit 19749 Reads einen Anteil von 3,57% an der Gesamtzahl der Reads von Region 1 hat. Die Test-Library hat eine durchschnittliche Leselänge von 355 bp. Die Größenverteilung der Test-Library-Reads zeigt, dass der

Anteil kurzer Sequenzen von bis zu 100 bp Länge mit 8,7% gering ist; die größte Fraktion mit 39,2% bilden Reads mit einer Länge von 401 bis 500 bp und der Anteil von Reads größer als 300 bp liegt bei 66,9% (vergleiche Abb. 68).

	TCAG-Wells Region 1	CATG-Wells Region 1
Raw Wells	951200	951200
Key Pass Wells	919627	7574
Passed Filter Wells	552883	6694
Total Bases	201381286	2574031
% Passed Filter	60,12	88,38
Lenght Average	365,19	384,59
Lenght Std Deviation	140,52	117,56
Longest Reads Lenght	2038	1137
Shortest Reads Lenght	40	41
Median Reads Lenght	409	409
Failed Dot	31802	91
Failed Mixed	85387	59
Failed Short Quality	249176	730
Failed Short Primer	379	0
% Dot + Mixed	12,74	1,98
% Short	27,14	9,64

Tab. 29: Run mit der Test-Library

Es sind die Ergebnisse von Region 1 dargestellt, die von der GSRUNBrowser-Software ausgegeben werden; die Spalte der TCAG-Wells zeigt die Library-Reads, die Spalte der CATG-Wells die Kontroll-DNA-Reads.

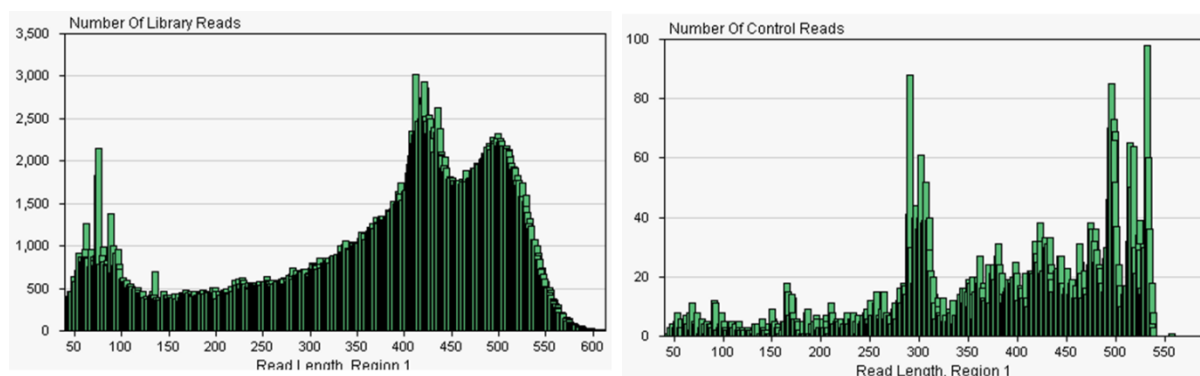


Abb. 65: Größenverteilung der Reads beim Run mit der Test-Library

Es werden die von der GSRUNBrowser-Software ausgegebenen Verteilungen der Library-Readlängen (links) und der Kontroll-DNA gezeigt (rechts). Es ist nur Region 1 dargestellt.

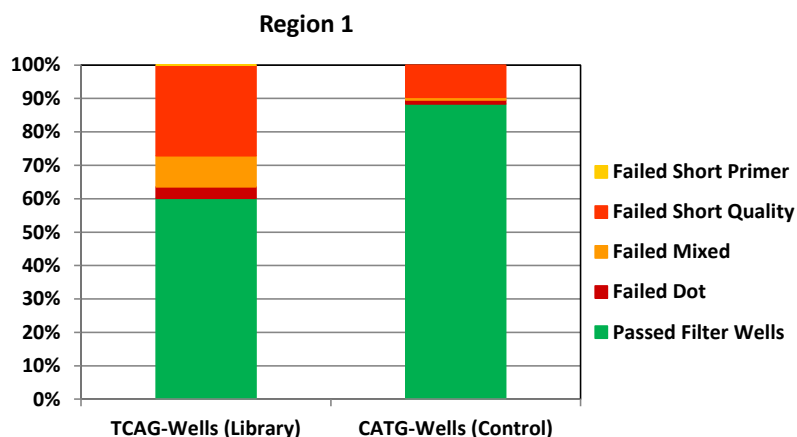


Abb. 66: Prozentuale Verteilung beim Run mit der Test-Library

Es wird das Verhältnis von Passed Filter Wells (grün) zu den Failed Wells auf Region 1 für Library-Reads (tcag-wells, links) und Kontroll-DNA-Reads (catg-wells, rechts) dargestellt.

Länge	100 bp			200 bp			300 bp			400 bp		
Match	100%	98%	95%	100%	98%	95%	100%	98%	95%	100%	98%	95%
Region 1	61,00	86,48	90,15	50,44	83,03	87,38	18,44	77,12	84,43	9,54	57,99	68,68

Tab. 30: Kontroll-DNA-Analyse beim Run mit der Test-Library

Es sind die von der GSRunBrowser-Software ausgegebenen Ergebnisse der Kontroll-DNA-Analyse für Region 1 dargestellt. Es wird der prozentuale Anteil der Kontroll-Bead-Reads angegeben, der zu je 100%, 98% und 95% mit den Kontroll-DNA-Sequenzen über eine Länge von 100 bp, 200 bp, 300 bp und 400 bp übereinstimmt (Match).

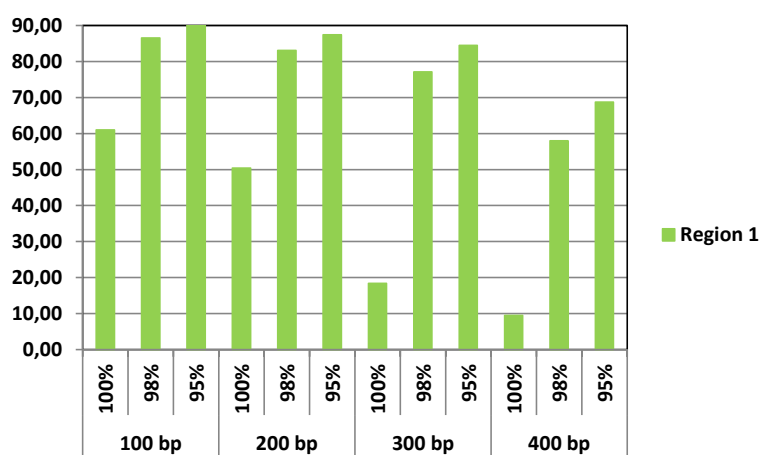


Abb. 67: Graphische Darstellung der Kontroll-DNA-Analyse

Es werden die Ergebnisse der Kontroll-DNA-Analyse von Region 1 aus Tab. 30 graphisch dargestellt.

MID	4	7	8	3	1	non-matching
Library	271-15	FR7	FR8	Test-Library	mge011	-
Read-Anzahl	464844	34559	27680	19749	1492	4559
Anteil in Prozent	84,08	6,25	5,01	3,57	0,27	0,82

Tab. 31: Zerlegung in die MIDs beim Run mit der Test-Library von Region 1

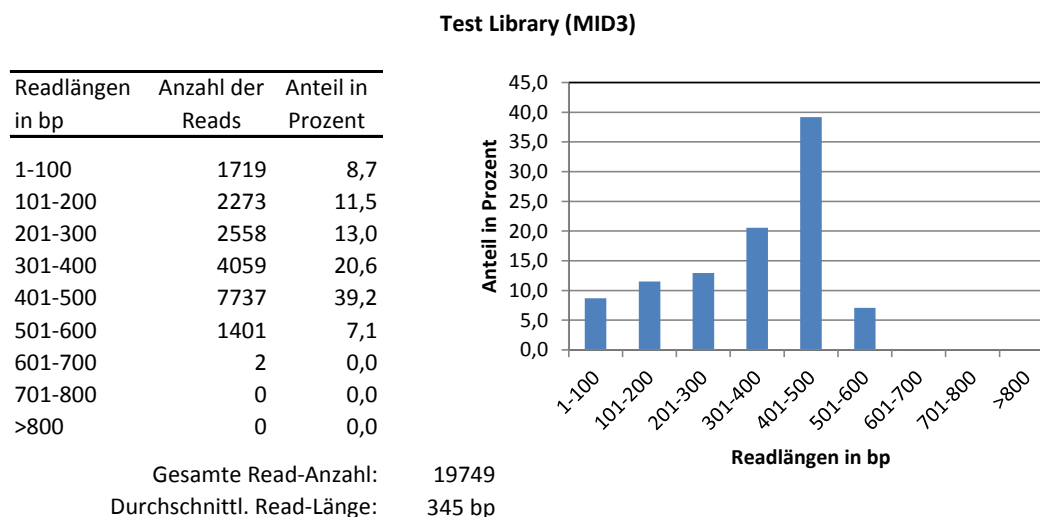


Abb. 68: Längenverteilung der Test-Library

Die Readlängen sind ohne die 10 bp lange MID-Sequenz angegeben.

Der gesamte Sequencing-Run als auch die Ergebnisse der Testlibrary stellen einen Erfolg dar. Die durchschnittliche Leselängen der Testlibrary und deren Größenverteilung sind gut und mit den Werten von genomischen Shotgun-Librarys vergleichbar. Es wurden nun 454-Läufe angesetzt, bei denen eine gesamte Region aus Librarys bestand, die mittels neuer Methodik erstellt wurden. Die Ergebnisse werden in den folgenden Kapiteln präsentiert.

3.3.2 Sequenzieren der aqf-Librarys mit der neuen Methodik

Aufgrund der positiven Ergebnisse der Test-Library sollten alle aqf-cDNAs mit der neuen Methodik erstellt und auf dem 454 sequenziert werden. Vorausgeplant war ein Run mit allen zwölf aqf-Librarys, bei vollständiger Beladung beider Regionen der PTP zu gleichen Teilen (vergleiche Seite 56 ff.).

3.3.2.1 Erstellen von aqf-cDNA-Librarys

Aus den aqf-cDNAs wurde mit der neu entwickelten Methode 454-Librarys hergestellt (Abschnitt 2.2.7.1 auf Seite 54 und 2.2.7.4 auf Seite 56). Nach der doppelten SPRI-Aufreinigung wurden die Konzentrationen der Librarys bestimmt (Tab. 32) und mit ihnen eine Test-PCR durchgeführt (Abb. 69).

Name	aqf3b	aqf4b	aqf5b	aqf6b	aqf9b	aqf10b	aqf15	aqf25	aqf35	aqf45	aqf55	aqf65
MID	20	21	22	23	24	25	20	21	22	23	24	25
Konzentration in ng/μl	37,4	38,4	35,2	8,4	28,5	29,7	35,6	30,4	37,2	30,6	26,9	28,3

Tab. 32: Konzentrationsbestimmung der aqf-Librarys

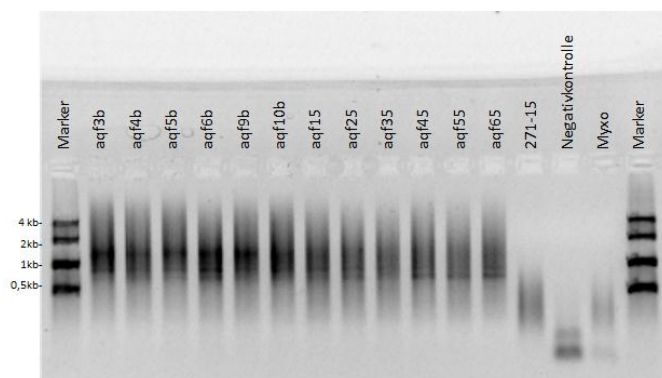


Abb. 69: Gelbild der Test-PCR mit den aqf-454-Librarys

Die Proben 271-15 und Myxo sind General-Librarys, die als Positiv-Kontrollen in der PCR mitgeführt wurden. Der Marker ist eine Eigenproduktion des MPI für molekulare Genetik.

Die Konzentrationen aller Librarys mit Ausnahme von aqf6b liegen in einem ähnlichen Bereich zwischen 26 bis 39 ng/μl. Auf dem Gelbild sind bei allen aqf-Library Amplifikations-Produkte in der erwarteten Größenordnung zwischen 0,5 bis 2 kb vorhanden; die Positivkontrollen (271-15 und Myxo) zeigen ebenfalls Produkte. Die Negativkontrolle ist nicht frei von Amplifikationsprodukt; dort sind zwei Banden zu sehen. Die untere, kleinere Bande kann unverbrauchter Primer sein (diese ist auch angedeutet bei der Probe „Myxo“ vorhanden), die obere Bande in der Negativkontrolle muss aber eine Kontamination oder ein unspezifisches Amplifikat sein.

3.3.2.2 Beadrecovery der ersten sechs aqf-Librarys

aqf3b, aqf4b, aqf5b, aqf6b, aqf9b und aqf10b wurden zu gleichen Teilen in einem emPCR-Cup angesetzt (vergleiche Tab. 16 auf Seite 56); anschließend folgte ein Beadrecovery (Tab. 33).

Librarys	zurückgewonnene Beads	Beadrecovery-Rate in Prozent	aufgereinigte Beads	Enrichment-Rate in Prozent
aqf3b				
aqf4b				
aqf5b				
aqf6b	30576322	87,4	4302500	14,1
aqf9b				
aqf10b				

Tab. 33: Beadrecovery der Librarys aqf3b bis aqf10b

Dargestellt sind die Messergebnisse nach Beadrecovery (zurückgewonnene Beads) und Enrichment (aufgereinigte Beads) und die daraus berechneten prozentualen Recovery-Rates. Es wird nur der Cup mit den aqf-Librarys gezeigt.

87,4% der verwendeten Beads wurden aus der Emulsion zurückgewonnen, womit das die höchste Ausbeute ist, die bis jetzt erreicht werden konnte. Die Enrichment-Rate von 14,1% liegt noch

unterhalb dem von Roche angegebenen Grenzbereich von 15-20% und spricht für gut sequenzierbare Beads. Die Anzahl an erhaltenen Beads (4,3 Mio.) reicht deutlich für die Beladung einer Region aus.

3.3.2.3 Sequencing-Run mit den ersten sechs aqf-Librarys

Auf Region 1 der PTP wurden die aufgereinigten Beads mit den Librarys aqf3b, aqf4b, aqf5b, aqf6b, aqgf9b und aqf10b beladen und auf dem 454 sequenziert -Region 2 bestand aus Beads mit General-Librarys. Tab. 34 und die Abbildungen auf den folgenden Seiten zeigen die Ergebnisse des Runs, Tab. 35 und Abb. 73 stellen die Kontroll-DNA-Ergebnisse dar. Tab. 36 führt die Zerlegung der Reads von Region 1 nach den MIDs auf, und Abb. 74 und Abb. 75 zeigen die Größenverteilungen der sechs aqf-Librarys.

Bei diesem Run kamen von Region 1 (aqf-Librarys) über 530000 Reads herunter (TCAG-Key), auf Region 2 über 600000, was einem Anteil von 67,2% Passed Filter Wells an den Key-Pass-Wells auf Region 1 entspricht (Tab. 34); auf Region 2 ist dieser sogar noch etwas größer mit 69,32%. Die durchschnittlichen Leseweiten betrugen auf Region 1 knapp 360 bp, auf Region 2 rund 410 bp, mit entsprechenden Größenverteilungen der Reads (Abb. 70). Beide Regionen stellen also ein erfolgreiches Sequencing dar, der gesamte Run liefert ein gutes Ergebnis.

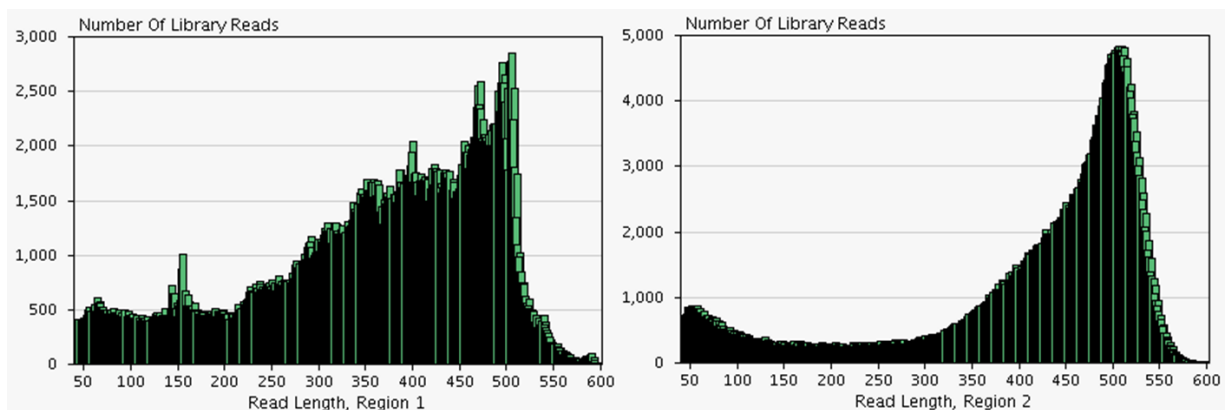
Es wurden dementsprechend wenig Reads durch die unterschiedlichen Filter herausortiert (Tab. 34 und Abb. 72); die größte Fraktion stellen dabei mit ca. 20% auf beiden Regionen diejenigen dar, die als Short herausgefiltert wurden. Der Anteil von Reads, die als Dot und Mixed entfernt wurde, ist mit rund 12% auf Region 1 bzw. 10% auf Region 2 klein (vergleiche Tab. 34 und Abb. 72). Die durchschnittliche Leselänge der Kontroll-DNA ist mit 385,56 bp auf Region 1 besser als auf Region 2 mit 373,52 bp (Tab. 34), wobei diese Leseweiten beider Regionen als gut zu werten sind; ihre Größenverteilungen sind entsprechend „rechtslastig“ (Abb. 71). Bei den Ausbeuten ist Region 1 mit 86,26% Passed Filter-Wells besser als Region 2 mit nur 73,23% bei den Kontroll-DNA-Reads; Region 2 liegt hier etwas unterhalb den Erwartungen, die man an einen so guten Run hat –es sollten über 80% der Kontroll-DNA-Reads als Passed-Filter-Wells die Filter im Signalprocessing passieren. Der Hauptunterschied zwischen Region 1 und Region 2 liegt in dem Anteil der Kontroll-DNA-Reads, die als Short herausortiert wurden, dieser ist auf Region 2 doppelt so groß wie auf zu Region 1 (vergleiche Tab. 34 und Abb. 72). Bei Übereinstimmungen (Match) von 100% über alle Längen zeigen die Kontroll-DNA-Reads von Region 2 bessere Ergebnisse (54,43% zu 46,52% von Region 1, 46,49% zu 19,75%, 23,33% zu 7,53% und 10,79% zu 3,9%), bei den Übereinstimmungen von 98% bzw. 95% ist es genau umgekehrt, hier zeigt Region 1 die besseren Ergebnisse (vergleiche Tab. 35 und Abb. 73).

Die Zerlegung in die verschiedenen MIDs bei Region 1 zeigt, dass die unterschiedlichen Librarys einigermaßen gleichmäßig vertreten sind, wobei aqf6b mit nur 9,3% einen etwas schwächeren und aqf10b mit 24,8% einen stärkeren Anteil an den Reads besitzt (Tab. 36). Aus der Aufschlüsselung der Größenverteilungen der sechs aqf-Librarys ist ersichtlich, dass bei allen Librarys der Anteil von kurzen Reads bis 100 bp am kleinsten ist und die größte Fraktion von Reads zwischen 401 und 500 bp gebildet wird (Abb. 74 und Abb. 75).

	TCAG-Wells (Library)			CATG-Wells (Control)		
	Region 1	Region 2	Total	Region 1	Region 2	Total
Raw Wells	848707	915481	1764188	848707	915481	1764188
Key Pass Wells	792609	874996	1667605	8375	8101	16476
Passed Filter Wells	532611	606523	1139134	7224	5932	13156
Total Bases	189999200	248598464	438597664	2785298	2215712	5001010
% Passed Filter	67,20	69,32	68,31	86,26	73,23	79,85
Lenght Average	359,73	409,88	385,03	385,56	373,52	387,6
Lenght Std Deviation	126,56	134,17		116,09	126,86	
Longest Reads Lenght	1086	1020	1086	544	1055	1055
Shortest Reads Lenght	40	40	40	40	40	40
Median Reads Lenght	383	460	427	412,5	404	432
Failed Dot	15002	26209	41211	21	21	42
Failed Mixed	79943	59685	139628	113	248	361
Failed Short Quality	159911	176303	336214	950	1813	2763
Failed Short Primer	843	184	1027	0	0	0
% Dot + Mixed	11,98	9,82	10,84	1,60	3,32	2,45
% Short	20,28	20,17	20,22	11,34	22,38	16,77

Tab. 34. Run der ersten sechs aqf-Librarys

Es werden die Ergebnisse des Runs dargestellt, die von der GSRUNBrowser-Software ausgegeben werden; links in der Tabelle befinden sich die der Library-DNA (tcag-wells) und rechts die der Kontroll-DNA (catg-wells). Region 1 bestand aus den Beads der aqf-Librarys.

**Abb. 70: Größenverteilung der Reads beim Run der ersten sechs aqf-Librarys**

Gezeigt sind die von der GSRUNBrowser-Software ausgegebenen Verteilungen der Readlängen von Sequenzen mit tcag-Key auf Region 1 (links –aqf-Librarys) und Region 2 (rechts).

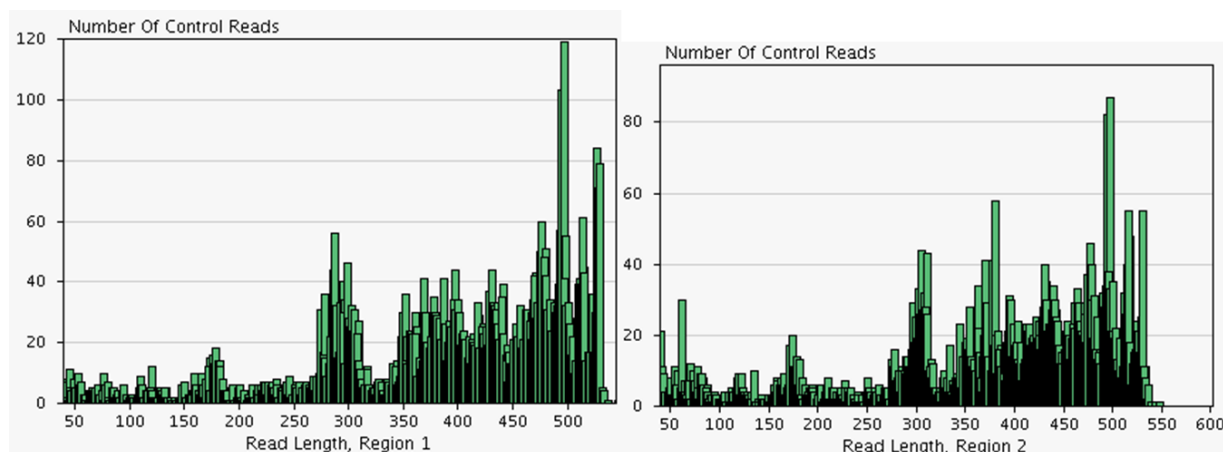


Abb. 71: Größenverteilung der Kontroll-DNA-Reads beim Run der ersten sechs aqf-Librarys

Es werden die von der GSRUNBrowser-Software ausgegebenen Verteilungen der Kontroll-DNA-Readlängen (Reads mit dem catg-Key) auf Region 1 (links) und Region 2 (rechts) dargestellt.

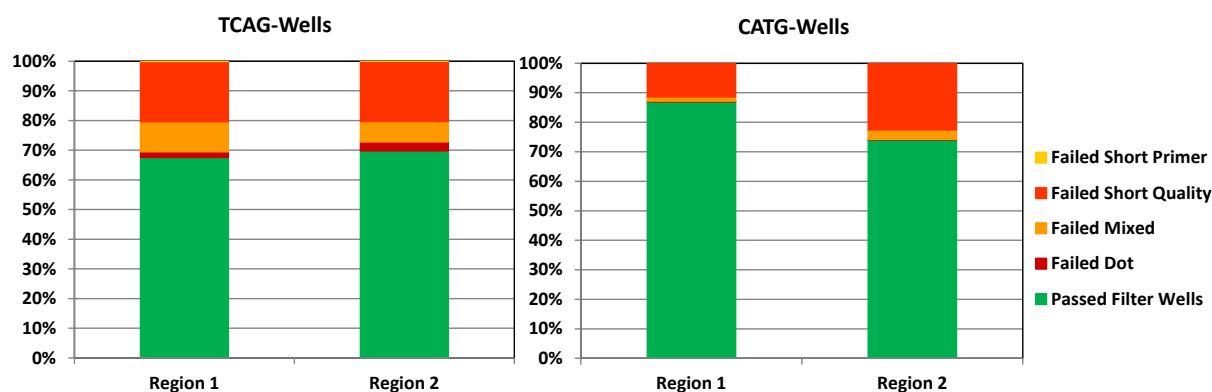


Abb. 72: Prozentuale Verteilung beim Run der ersten sechs aqf-Librarys

Gezeigt wird das Verhältnis auf beiden Regionen von Passed Filter Wells (grün) zu Failed Wells, für Library-Reads (tcag-wells) und Kontroll-DNA-Reads (catg-wells). Die aqf-Librarys waren auf Region 1 geladen.

Länge	100 bp			200 bp			300 bp			400 bp		
Match	100%	98%	95%	100%	98%	95%	100%	98%	95%	100%	98%	95%
Region 1	46,52	78,70	84,84	19,57	73,34	80,37	7,53	66,95	78,01	3,90	53,06	65,68
Region 2	54,43	67,31	72,41	46,49	62,72	68,57	23,23	58,15	65,67	10,79	47,64	56,86

Tab. 35: Kontroll-DNA-Analyse beim Run der ersten sechs aqf-Librarys

Es werden die von der GSRUNBrowser-Software ausgegebenen Ergebnisse der Kontroll-DNA-Analyse gezeigt. Für beide Regionen des Runs wird der prozentuale Anteil der Kontroll-Bead-Reads dargestellt, der zu je 100%, 98% und 95% mit den Kontroll-DNA-Sequenzen über eine Länge von 100 bp, 200 bp, 300 bp und 400 bp übereinstimmt (Match).

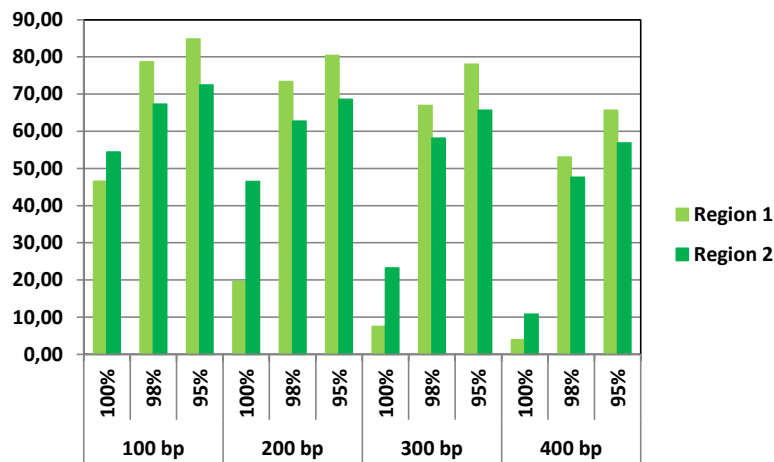


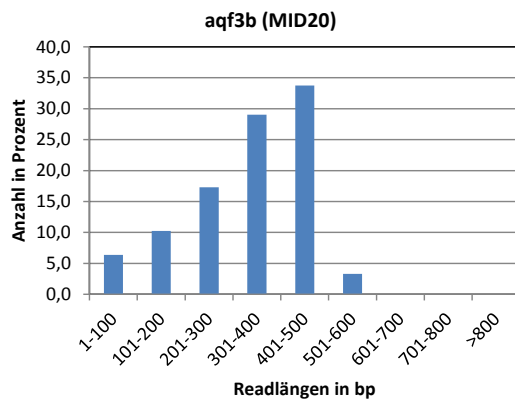
Abb. 73: Graphische Darstellung der Kontroll-DNA-Analyse

Es werden die Ergebnisse der Kontroll-DNA-Analyse aus Tab. 35 graphisch dargestellt.

MID	20	21	22	23	24	25	non-matching
Library	aqf3b	aqf4b	aqf5b	aqf6b	aqf9b	aqf10b	-
Read-Anzahl	85321	90995	70321	49558	102795	132294	1327
Anteil in Prozent	16,02	17,08	13,20	9,30	19,30	24,84	0,25

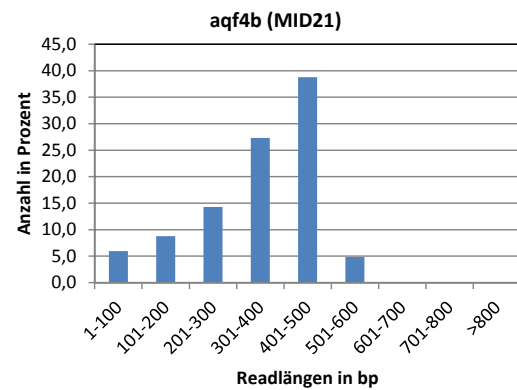
Tab. 36: Zerlegung in die MIDs beim Run der ersten sechs aqf-Librarys

Es wird nur Region 1 angegeben, auf der die aqf-Librarys geladen waren.



Gesamte Read-Anzahl: 85321
Durchschnittl. Read-Länge: 336 bp

Readlängen in bp	Anzahl der Reads	Anteil in Prozent
1-100	5433	6,4
101-200	8734	10,2
201-300	14769	17,3
301-400	24757	29,0
401-500	28810	33,8
501-600	2815	3,3
601-700	1	0,0
701-800	0	0,0
>800	2	0,0

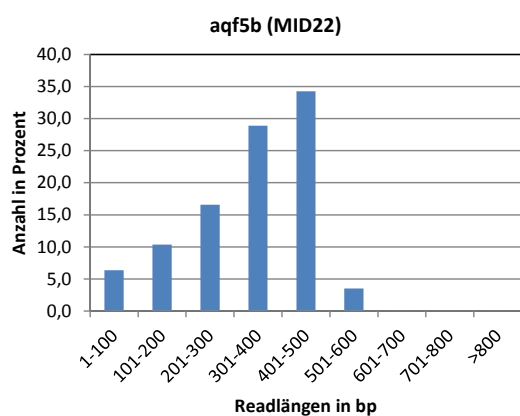


Gesamte Read-Anzahl: 90995
Durchschnittl. Read-Länge: 352 bp

Readlängen in bp	Anzahl der Reads	Anteil in Prozent
1-100	5444	6,0
101-200	7990	8,8
201-300	12987	14,3
301-400	24841	27,3
401-500	35296	38,8
501-600	4424	4,9
601-700	9	0,0
701-800	1	0,0
>800	2	0,0

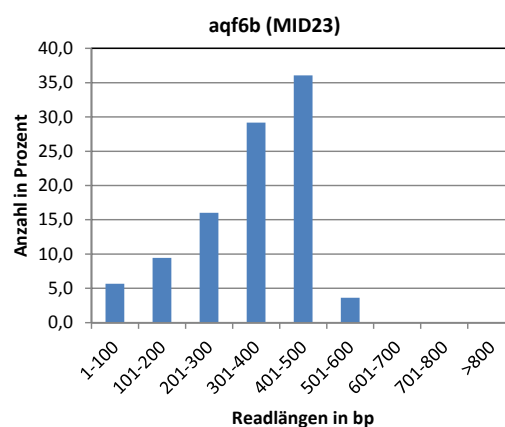
Abb. 74: Größenverteilungen von aqf3b und aqf4b

Bei der Angabe der durchschnittlichen Read-Länge wurde das 10 bp lange MID abgezogen.



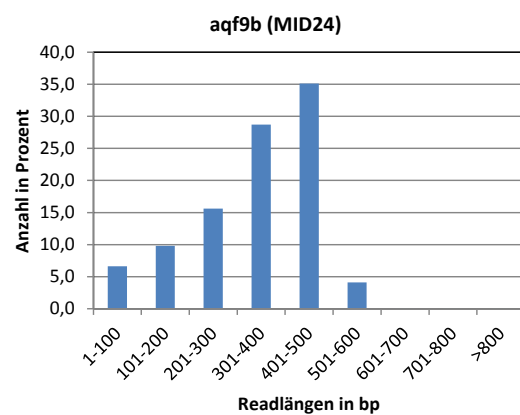
Gesamte Read-Anzahl: 70321
Durchschnittl. Read-Länge: 338 bp

Readlängen in bp	Anzahl der Reads	Anteil in Prozent
1-100	4460	6,3
101-200	7292	10,4
201-300	11658	16,6
301-400	20323	28,9
401-500	24111	34,3
501-600	2470	3,5
601-700	4	0,0
701-800	1	0,0
>800	2	0,0



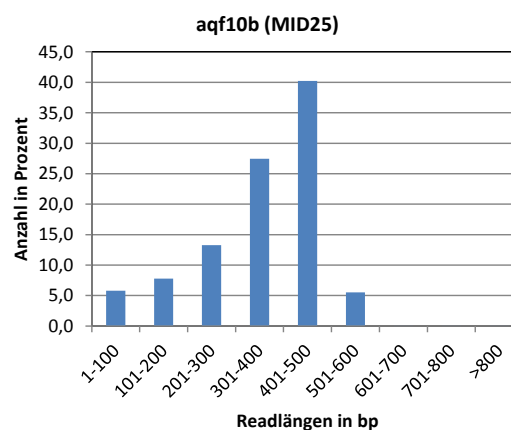
Gesamte Read-Anzahl: 49558
Durchschnittl. Read-Länge: 345 bp

Readlängen in bp	Anzahl der Reads	Anteil in Prozent
1-100	2803	5,7
101-200	4677	9,4
201-300	7934	16,0
301-400	14459	29,2
401-500	17883	36,1
501-600	1795	3,6
601-700	5	0,0
701-800	0	0,0
>800	2	0,0



Gesamte Read-Anzahl: 102795
Durchschnittl. Read-Länge: 341 bp

Readlängen in bp	Anzahl der Reads	Anteil in Prozent
1-100	6800	6,6
101-200	10081	9,8
201-300	16053	15,6
301-400	29519	28,7
401-500	36094	35,1
501-600	4233	4,1
601-700	8	0,0
701-800	4	0,0
>800	3	0,0



Gesamte Read-Anzahl: 132294
Durchschnittl. Read-Länge: 358 bp

Readlängen in bp	Anzahl der Reads	Anteil in Prozent
1-100	7634	5,8
101-200	10276	7,8
201-300	17576	13,3
301-400	36297	27,4
401-500	53211	40,2
501-600	7277	5,5
601-700	21	0,0
701-800	1	0,0
>800	1	0,0

Abb. 75: Größenverteilungen von aqf5b, aqf6b, aqf9b und aqf10b

Bei der Angabe der durchschnittlichen Read-Länge wurde das 10 bp lange MID abgezogen.

Da Ausbeuten, durchschnittliche Leselängen und Größenverteilungen der Librarys auf Region 1 des Laufs gut sind, wurde ein weiterer Sequenzier-Lauf mit den weiteren aqf-Librarys angesetzt, welche nach der neuen Methodik hergestellt wurden.

3.3.2.4 Beadrecovery der weiteren sechs aqf-Librarys

In einem emPCR-Cup wurden zu gleichen Teilen aqf15, aqf25, aqf35, aqf45, aqf55 und aqf65 angesetzt (vergleiche Tab. 17 auf Seite 57) und nach der Amplifikation ein Beadrecovery durchgeführt (Tab. 37).

Mit einer Beadrecovery-Rate von 80% liegt man etwas unterhalb den von Roche vorgegebenen 85%. Die Enrichment-Rate von 13,3% liegt unterhalb des angegebenen Grenzbereiches von 15-20%. Die erhaltenen Beads sollten also gute Sequenzier-Ergebnisse liefern; mit über 3,7 Mio. stehen auch genug Beads für die Beladung einer PTP-Region zur Verfügung.

Librarys	zurückgewonnene Beads	Beadrecovery-Rate in Prozent	aufgereinigte Beads	Enrichment-Rate in Prozent
aqf15				
aqf25				
aqf35	27984655	80,0	3732440	13,3
aqf45				
aqf55				
aqf65				

Tab. 37: Beadrecovery der Librarys aqf15 bis aqf65

Es werden die Messergebnisse nach Beadrecovery (zurückgewonnene Beads) und Enrichment (aufgereinigte Beads) und die daraus berechneten prozentualen Recovery-Rates gezeigt. Es wird nur der Cup mit den aqf-Librarys dargestellt.

3.3.2.5 Sequencing-Run mit den weiteren sechs aqf-Librarys

Die Beads mit den Librarys aqf15, aqf25, aqf35, aqf45, aqf55 und aqf65 wurden auf PTP-Region 1 geladen und sequenziert. Die PTP-Region 2 wurde mit General-Library-Beads beladen. In Tab. 38 und den Abbildungen auf den folgenden Seiten sind die Ergebnisse des Runs dargestellt, Tab. 39 und Abb. 79 zeigen die Kontroll-DNA-Ergebnisse. In Tab. 40 ist die Zerlegung der Reads von Region 1 nach den MIDs angegeben, und Abb. 80 und Abb. 81 zeigen die Größenverteilungen der sechs aqf-Librarys.

Bei diesem Run lieferte Region 1 (aqf Librarys) über 478000 Reads mit TCAG-Key, Region 2 mehr als 660000 (Tab. 38). Das ergibt einen Anteil von 62,27% Passed Filter Wells auf Region 1 an den Keypass-Wells, bei Region 2 sind es 69,91%. Damit liegt auch dieser Run oberhalb der von Roche angegebenen Untergrenze von 400000 Reads pro Region für einen guten Run (mit Shotgun-Librarys). Die durchschnittlichen Leseweiten der Library-Reads (TCAG-Wells) von 364 bp (Region 1) bzw. 410 bp (Region 2) mit entsprechenden rechtslastigen Größenverteilungen (Abb. 76), liegen ebenfalls über der angegebenen Untergrenze von 350 bp. Auch dieser Sequencing-Run zeigt also auf beiden Regionen gute Ergebnisse und kann als Erfolg gewertet werden.

Der Anteil an Reads, der durch die verschiedenen Filter entfernt wurde, ist dementsprechend gering. Auf Region 1 wurden 23% der Library-Reads als Short heraussortiert, auf Region 2 18%, was auch bei beiden Regionen die größten Fraktionen bei den Failed Reads sind (Abb. 78) -der Dot- und Mixed-Anteil liegt auf beiden Regionen unterhalb von 15%.

Mit 410 bp auf Region 1 und 393 bp auf Region 2 sind die durchschnittlichen Leseweiten der Kontroll-DNA hoch (CATG-Wells, Tab. 38), ihre Größenverteilung fällt entsprechend aus (Abb. 77). Die Ausbeuten der Kontroll-DNA sind auf beiden Regionen ebenfalls hoch, beide Regionen weisen Passed-Filter-Werte von über 94% auf, der Anteil an Kontroll-DNA-Reads, die von den Filtern also Short, Dot oder Mixed entfernt wurden, ist sehr klein und liegt insgesamt unter 6% (Abb. 78 und Tab. 38). Die Kontroll-DNA-Analyse zeigt (Tab. 39 und Abb. 79), dass beide Regionen fast identische Ergebnisse bei den 98%igen und 95%igen Übereinstimmungen über alle Längen zeigen; bei den 100%igen Übereinstimmungen weichen die Anteile der Kontroll-DNA-Reads auf Region 1 bei 100 bp und 200 bp Länge etwas stärker von denen von Region 2 ab und sind um 16% bis 17% niedriger.

Die Zerlegung der Reads auf Region 1 in die unterschiedlichen MIDs zeigt, dass die sechs Librarys relativ gleichmäßig verteilt sind, die Anteile bewegen sich zwischen rund 12% und 21% (Tab. 40). Bei allen Librarys ist der Anteil an kurzen Reads bis 100 bp am geringsten und auch hier wird die größte Fraktion von Reads zwischen 401 und 500 bp gebildet mit Anteilen von jeweils über 37% (Abb. 80 und Abb. 81).

	TCAG-Wells (Library)			CATG-Wells (Control)		
	Region 1	Region 2	Total	Region 1	Region 2	Total
Raw Wells	826045	988510	1814555	826045	988510	1814555
Key Pass Wells	768661	944650	1713311	16951	9109	26060
Passed Filter Wells	478622	660364	1138986	16041	8589	24630
Total Bases	174333216	270912067	445245283	6573156	3375342	9948498
% Passed Filter	62,27	69,91	66,48	94,63	94,29	94,51
Lenght Average	364,24	410,25	390,91	409,77	392,98	403,92
Lenght Std Deviation	129,96	115,47		97,58	105,12	
Longest Reads Lenght	1194	1196	1196	539	584	584
Shortest Reads Lenght	40	40	40	40	40	40
Median Reads Lenght	398	444	427	434	411	428
Failed Dot	22766	29658	52424	49	13	62
Failed Mixed	85772	79111	164883	71	31	102
Failed Short Quality	176410	170355	346765	705	429	1134
Failed Short Primer	772	139	911	0	0	0
% Dot + Mixed	14,12	11,51	12,68	0,71	0,48	0,63
% Short	23,05	18,05	20,29	4,16	4,71	4,35

Tab. 38: Run mit den weiteren sechs aqf-Librarys

Links in der Tabelle befinden sich die Ergebnisse der Library-DNA (tcag-wells) des Runs und rechts die der Kontroll-DNA (catg-wells), die von der GSRUNBrowser-Software ausgegeben werden. Auf Region 1 waren die aqf-Library-Beads geladen.

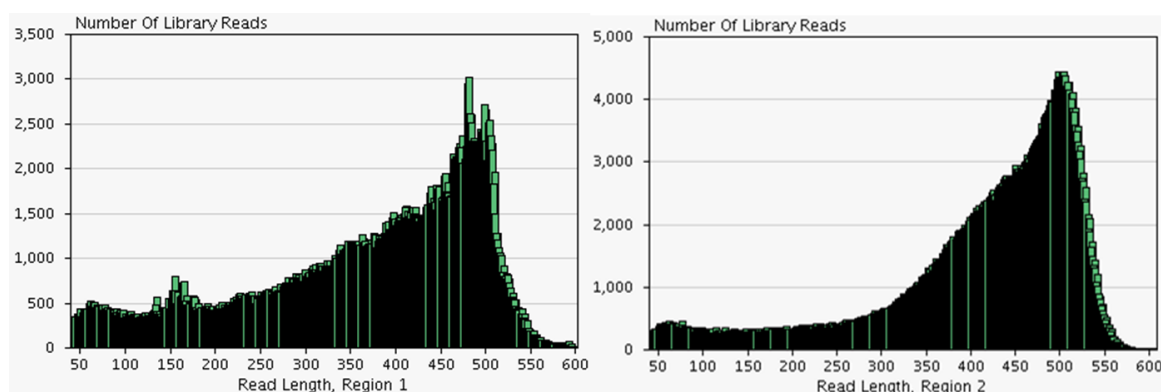


Abb. 76: Größenverteilung der Reads beim Run mit den weiteren aqf-Librarys

Es werden die Verteilungen der von der GSRUNBrowser-Software ausgegebenen Readlängen mit tcag-Key (Library-DNA) für beide Regionen dargestellt. Region 1 (links) bestand aus den aqf-Beads.

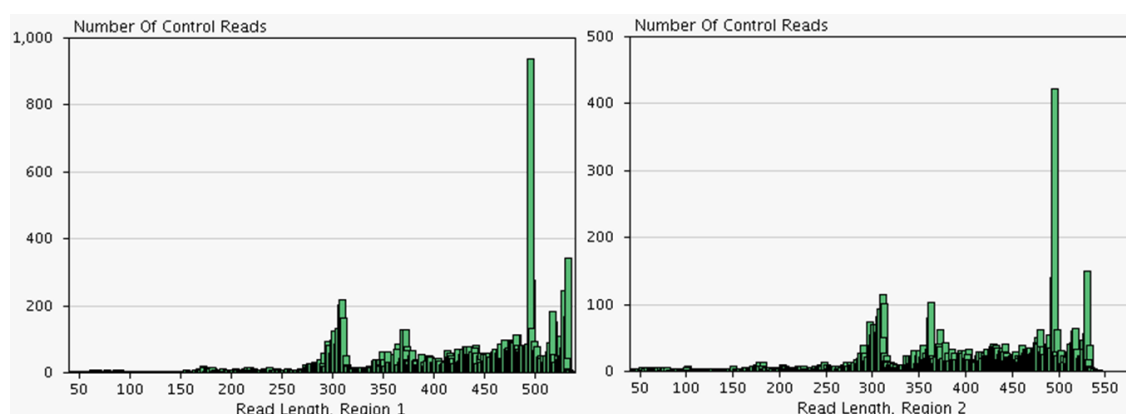


Abb. 77: Größenverteilung der Kontroll-DNA-Reads beim Run mit den weiteren aqf-Librarys

Dargestellt sind die von der GSRUNBrowser-Software ausgegebenen Verteilungen der Kontroll-DNA-Readlängen (Reads mit dem catg-Key) auf beiden Regionen.

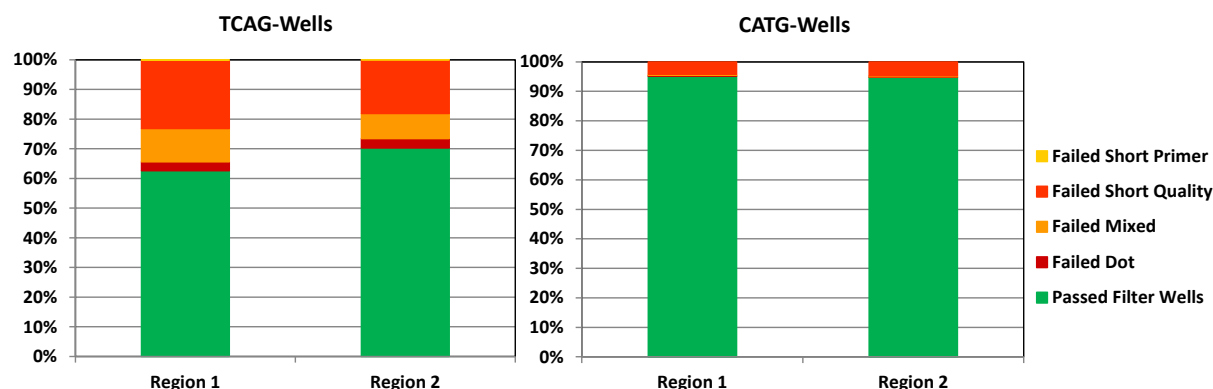


Abb. 78: Prozentuale Verteilung beim Run mit den weiteren aqf-Librarys

Angegeben ist das Verhältnis von Passed Filter Wells (grün) zu Failed Wells auf beide Regionen für die Library-Reads (tcag-wells) und Kontroll-DNA-Reads (catg-wells). Die aqf-Librarys waren auf Region 1 geladen.

Länge	100 bp			200 bp			300 bp			400 bp		
	100%	98%	95%	100%	98%	95%	100%	98%	95%	100%	98%	95%
Region 1	58,03	89,95	93,56	47,44	88,97	92,68	23,98	83,25	90,44	12,17	62,80	76,23
Region 2	74,73	90,73	93,47	64,98	87,96	91,80	30,08	82,47	88,92	14,57	62,66	74,46

Tab. 39: Kontroll-DNA-Analyse beim Run mit den weiteren aqf-Librarys

Es werden die von der GSRUNBrowser-Software ausgegebenen Ergebnisse der Kontroll-DNA-Analyse dargestellt. Für beide Regionen des Runs wird der prozentuale Anteil der Kontroll-Bead-Reads gezeigt, der zu je 100%, 98% und 95% mit den Kontroll-DNA-Sequenzen über eine Länge von 100 bp, 200 bp, 300 bp und 400 bp übereinstimmt (Match).

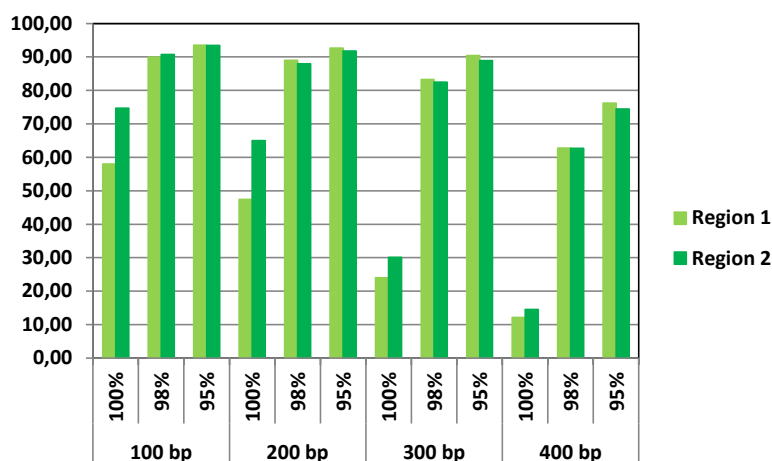


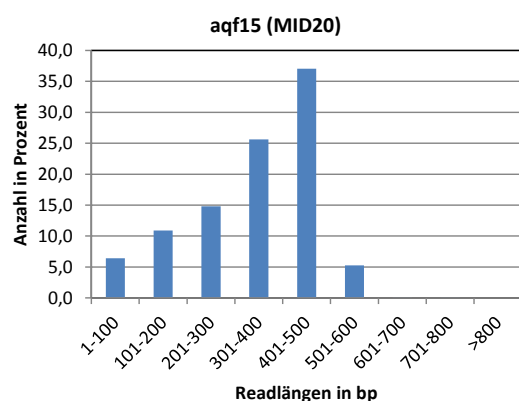
Abb. 79: Graphische Darstellung der Kontroll-DNA-Analyse

Es werden die Ergebnisse aus Tab. 39 graphisch dargestellt.

MID	20	21	22	23	24	25	non-matching
Library	aqf15	aqf25	aqf35	aqf45	aqf55	aqf65	-
Read-Anzahl	61407	59285	70673	102334	101028	82762	1133
Anteil in Prozent	12,83	12,39	14,77	21,38	21,11	17,29	0,24

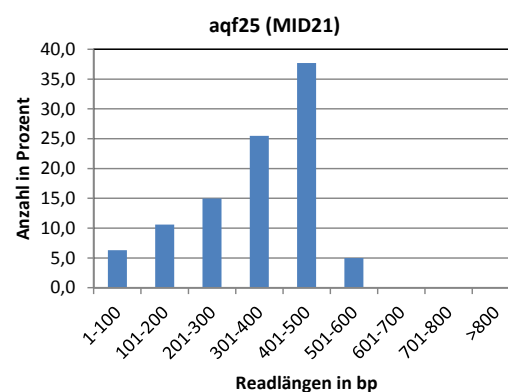
Tab. 40: Zerlegung in die MIDs beim Run der weiteren aqf-Librarys

Angegeben ist das Ergebnis für Region 1, auf der die aqf-Librarys geladen waren.



Gesamte Read-Anzahl: 61407
Durchschnittl. Read-Länge: 344 bp

Readlängen in bp	Anzahl der Reads	Anteil in Prozent
1-100	3922	6,4
101-200	6695	10,9
201-300	9082	14,8
301-400	15721	25,6
401-500	22758	37,1
501-600	3219	5,2
601-700	8	0,0
701-800	2	0,0
>800	0	0,0

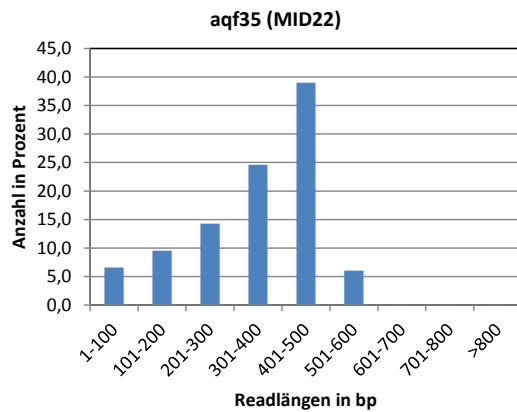


Gesamte Read-Anzahl: 59285
Durchschnittl. Read-Länge: 345 bp

Readlängen in bp	Anzahl der Reads	Anteil in Prozent
1-100	3734	6,3
101-200	6269	10,6
201-300	8882	15,0
301-400	15107	25,5
401-500	22337	37,7
501-600	2945	5,0
601-700	8	0,0
701-800	2	0,0
>800	1	0,0

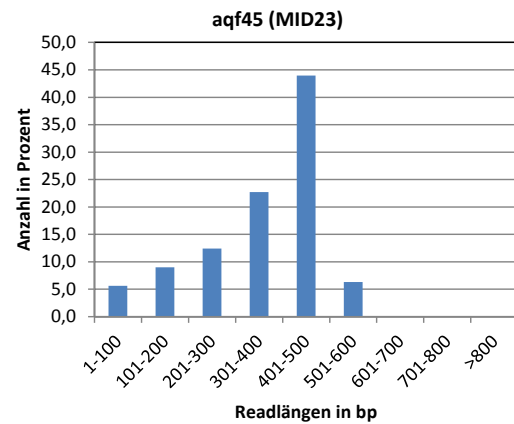
Abb. 80: Größenverteilungen von aqf15 und aqf25

Bei der Angabe der durchschnittlichen Read-Länge wurde das 10 bp lange MID abgezogen.



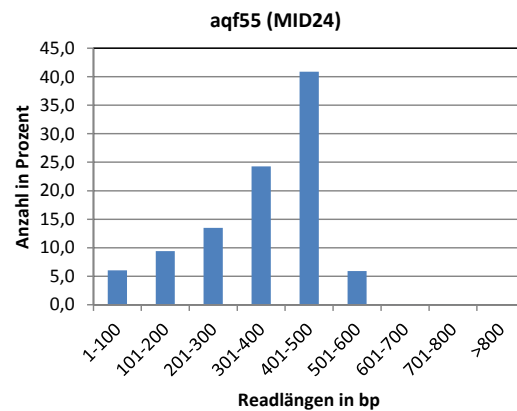
Gesamte Read-Anzahl: 70673
Durchschnittl. Read-Länge: 351 bp

Readlängen in bp	Anzahl der Reads	Anteil in Prozent
1-100	4631	6,6
101-200	6727	9,5
201-300	10083	14,3
301-400	17392	24,6
401-500	27566	39,0
501-600	4264	6,0
601-700	6	0,0
701-800	2	0,0
>800	2	0,0



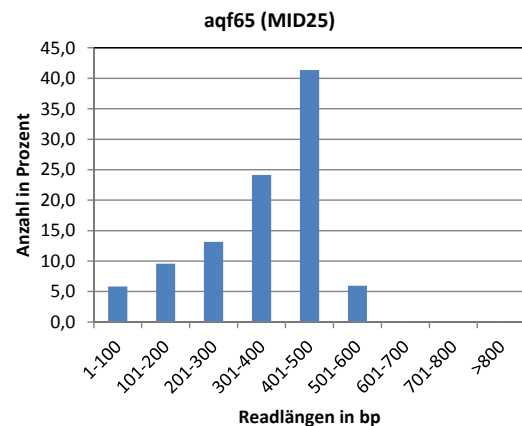
Gesamte Read-Anzahl: 102334
Durchschnittl. Read-Länge: 363 bp

Readlängen in bp	Anzahl der Reads	Anteil in Prozent
1-100	5741	5,6
101-200	9196	9,0
201-300	12693	12,4
301-400	23218	22,7
401-500	45001	44,0
501-600	6471	6,3
601-700	11	0,0
701-800	0	0,0
>800	3	0,0



Gesamte Read-Anzahl: 101028
Durchschnittl. Read-Länge: 355 bp

Readlängen in bp	Anzahl der Reads	Anteil in Prozent
1-100	6103	6,0
101-200	9481	9,4
201-300	13633	13,5
301-400	24533	24,3
401-500	41297	40,9
501-600	5966	5,9
601-700	10	0,0
701-800	5	0,0
>800	0	0,0



Gesamte Read-Anzahl: 82762
Durchschnittl. Read-Länge: 357 bp

Readlängen in bp	Anzahl der Reads	Anteil in Prozent
1-100	4832	5,8
101-200	7912	9,6
201-300	10882	13,1
301-400	19962	24,1
401-500	34232	41,4
501-600	4934	6,0
601-700	6	0,0
701-800	0	0,0
>800	2	0,0

Abb. 81: Größenverteilungen von aqf35, aqf45, aqf55, aqf65

Bei der Angabe der durchschnittlichen Read-Länge wurde das 10 bp lange MID abgezogen.

Die Sequenzierung der sechs weiteren aqf-Librarys ist ebenfalls ein Erfolg. Ausbeuten, durchschnittliche Leselängen und die Größenverteilungen der Librarys sind gut und mit genomischen Shotgun-Librarys vergleichbar.

Es konnten so weitere Librarys anderer Projekte sequenziert werden, die wie die aqf-cDNAs zuerst mittels MINT-Kit von der AG Kube in cDNA umgewandelt und aus denen dann mittels neuer Methodik 454 Librarys erstellt wurden. So konnte nach Herstellen einer 454-Library dem neuen Protokoll folgend aus haustorialer cDNA von *Golovinomyces orontii* im Sequencing rund 881000 Reads verteilt auf mehrere Runs erhalten werden (306 Mb Gesamtdatenmenge), was in einer Publikation mündete [123].

3.3.3 Erweiterung auf andere Templates

In den folgenden Abschnitten werden die Ergebnisse von Library-Erstellung und Sequenzierung anhand von zwei weiteren Template-DNAs demonstriert. Es sollte gezeigt werden, inwieweit sich die neue Methodik zur Library-Erstellung für DNAs eignet, die nicht oder nur teilweise mit dem MINT-Kit erstellt wurden und sie somit auf andere Bereiche ausgeweitet werden kann. Als erstes Beispiel dient ein Template, das keinerlei MINT-Sequenzen trägt und ursprünglich eine Illumina-Library darstellt, als zweites Beispiel wird die neue Methodik zur Antikörper-cDNA-Sequenzierung herangezogen.

3.3.3.1 Sequenzierung einer Illumina-Library

In diesem Abschnitt wird geprüft, ob sich die neue Methodik auch für Template-DNAs eignet, die keine über das MINT-Kit eingeführten SfiI-Schnittstellen besitzen. Das Prinzip der Vorgehensweise wird in Kapitel 2.2.7.9 gezeigt.

An 1450 ng einer 250 bp langen DNA (Bezeichnung OCon), die ursprünglich eine fertige Illumina-cDNA-Library darstellte, wurden mittels PCR SfiI-Schnittstellen heranamplifiziert und daraus dann nach der neuen Methodik eine 454-Library hergestellt (vergleiche Seite 58). Die Konzentration der so erhaltenen 454-OCon-Library wurde auf 9,4 ng/µl gemessen, was bei einem Gesamtvolumen von 100 µl eine Gesamtmenge von 940 ng Library-DNA ergibt. Diese Library wurde in einer emPCR mit anderen 454-Librarys (welche als Leitproben fungierten) angesetzt und die erhaltenen Beads auf eine PTP-Region für das Sequenzieren mit dem 454 geladen. In Tab. 41 und den Abbildungen auf den folgenden Seiten sind die Ergebnisse des Runs dargestellt; Tab. 42 und Abb. 84 zeigen eine Analyse der Kontroll-DNA und Abb. 85 zeigt die Längenverteilung der OCon-Library.

Region 1 erzeugte über 490000 Reads (Sequenzen mit TCAG-Key) mit einem Anteil von fast 60% (Passed Filter Wells) an den Key-Pass-Wells (Tab. 41 auf Seite 108). Die durchschnittliche Leseweite liegt bei 335 bp. Ursache dafür sind die Librarys, die neben der OCon-Library auf die PTP geladen waren; diese fielen in vorangegangenen Läufen schon kurz aus bzw. waren tatsächlich nicht länger

(kurze cDNAs, Ergebnisse nicht angegeben). Bei den Failed Reads ist der Anteil der „Short“-Reads mit 23,61 am größten (Abb. 83), der Anteil der Herausgefilterten Reads liegt mit fast 40% etwas höher, als man es für einen sehr guten Run erwarten würde (30-15%).

Die Kontroll-DNA-Reads zeigen mit einer durchschnittlichen Leselänge von 415 bp, entsprechender Größenverteilung (Abb. 82) und Ausbeuten von 85% Passed Filter Wells gute Ergebnisse; auch die Qualität der erzeugten Kontroll-DNA-Reads zeigt hohe Anteile bei den verschiedenen Übereinstimmungen (100%, 98% und 95%) über alle Längen (Tab. 42 und Abb. 84).

Die Analyse der Längenverteilung der OCon-Library ergibt, dass sie eine durchschnittliche Leselänge von 320 bp (bzw. 330 bp mit MID) besitzt. Laut den Angaben des Auftraggebers besaß die ursprüngliche Illumina-Library eine durchschnittliche Länge von ca. 250 bp. Die mit dem 454 erreichten längeren durchschnittlichen Leseweiten sind mit den hinzugefügten Sequenzen zu erklären (454-Adapter, MID, Sfil-Schnittstellen). Die größte Fraktion mit einem Anteil von 28,8% bilden dabei Reads mit einer Länge zwischen 301 und 400 bp, der Großteil der Reads liegt zwischen 301 bis 500 bp (57,2% insgesamt) und die Fraktion an Reads, die bis 100 bp lang sind, ist mit 10,6% die kleinste (Abb. 85).

	TCAG-Wells Region 1	CATG-Wells Region 1
Raw Wells	856898	856898
Key Pass Wells	822414	8630
Passed Filter Wells	491465	7372
Total Bases	164303972	3069321
% Passed Filter	59,76	85,42
Lenght Average	335,03	415,78
Lenght Std Deviation	143,83	116,63
Longest Reads Lenght	1591	851
Shortest Reads Lenght	40	46
Median Reads Lenght	364	460
Failed Dot	69704	58
Failed Mixed	67081	141
Failed Short Quality	189485	1059
Failed Short Primer	4679	0
% Dot + Mixed	16,63	2,31
% Short	23,61	12,27

Tab. 41: Runs mit der OCon-Library

Dargestellt sind die von der GSRUNBrowser-Software ausgegebenen Ergebnisse von Region 1, auf die die OCon-Library geladen war. Die linke Spalte (TCAG-Wells) zeigt die Library-Reads, die rechte (CATG-Wells) die Kontroll-DNA-Reads.

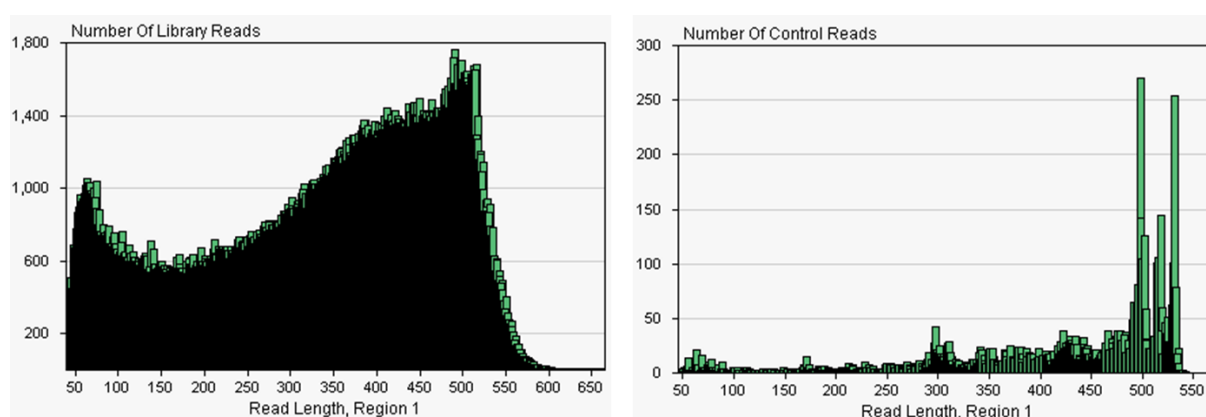


Abb. 82: Größenverteilung der Reads beim Run mit der OCon-Library

Es sind die Verteilungen der Library-Readlängen (links) und der Kontroll-DNA-Reads (rechts) auf Region 1 angezeigt, wie sie von der GSRUNBrowser-Software ausgegeben werden.

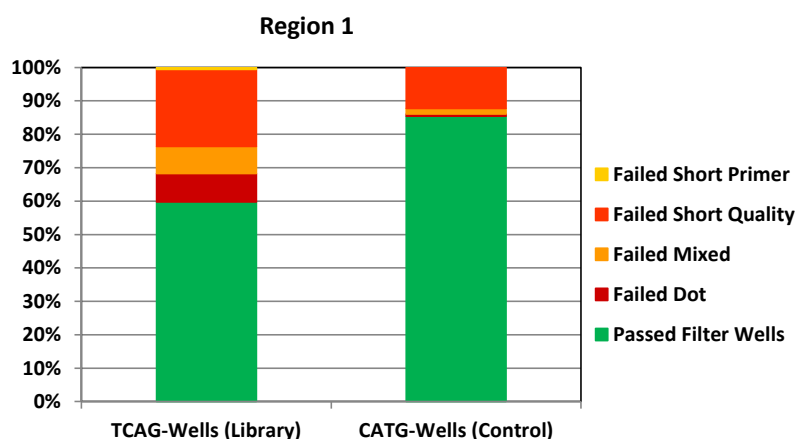


Abb. 83: Prozentuale Verteilung beim Run mit der OCon-Library

Für Region 1 wird das Verhältnis von Passed Filter Wells (grün) zu den Failed Wells dargestellt. Die Library-Reads sind links und die Kontroll-DNA-Reads rechts abgebildet.

Länge	100 bp			200 bp			300 bp			400 bp		
Match	100%	98%	95%	100%	98%	95%	100%	98%	95%	100%	98%	95%
Region 1	65,32	82,80	87,03	54,50	80,48	85,24	27,96	75,59	82,92	13,33	64,48	76,88

Tab. 42: Kontroll-DNA-Analyse beim Run mit der OCon-Library

Es sind für Region 1 die von der GSRUNBrowser-Software ausgegebenen Ergebnisse der Kontroll-DNA-Analyse dargestellt. Angezeigt werden die prozentualen Anteile der Kontroll-Bead-Reads, die zu je 100%, 98% und 95% mit den Kontroll-DNA-Sequenzen über Längen von 100 bp, 200 bp, 300 bp und 400 bp übereinstimmen (Match).

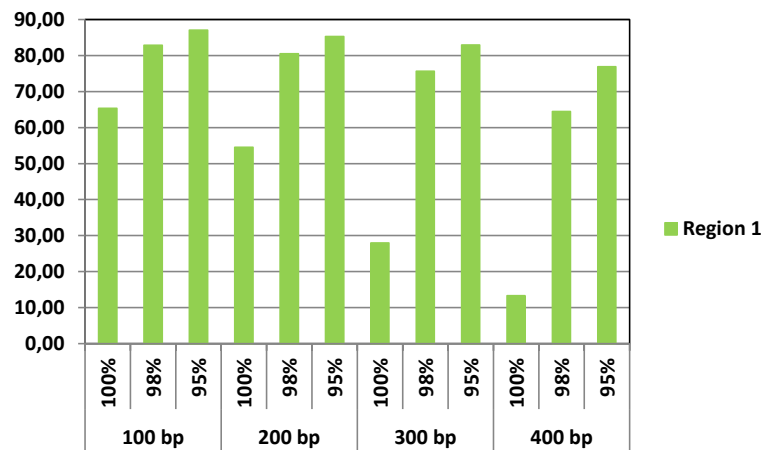
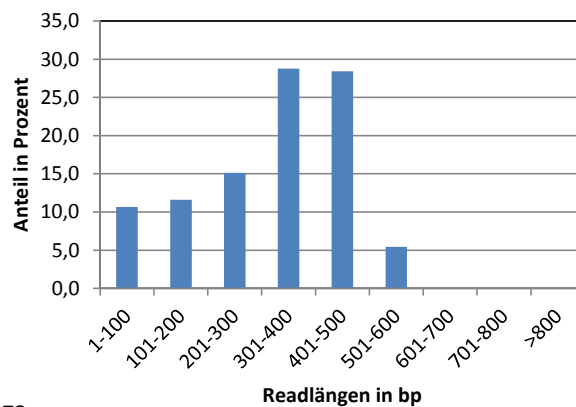


Abb. 84: Graphische Darstellung der Kontroll-DNA-Analyse

Es werden die Ergebnisse aus Tab. 42 graphisch dargestellt.

OCon (MID26)

Readlängen in bp	Anzahl der Reads	Anteil in Prozent
1-100	11856	10,6
101-200	12914	11,6
201-300	16843	15,1
301-400	32029	28,8
401-500	31664	28,4
501-600	6060	5,4
601-700	4	0,0
701-800	1	0,0
>800	2	0,0



Gesamte Read-Anzahl: 111373
Durchschnittl. Read-Länge: 320

Abb. 85: Längenverteilungen der OCon-Library

Die Readlängen sind ohne die 10 bp lange MID-Sequenz angegeben.

Sowohl der Run, auf dem die OCon-Library geladen war, als auch die Sequenzierungsergebnisse der OCon-Library stellen einen Erfolg dar. Deren durchschnittliche Leseweiten zeigen, dass zum Großteil die Sequenzierung über die gesamte Länge der Illumina-Library möglich war. Das zeigt außerdem, dass die Erweiterung der Methodik auf andere Templates funktioniert und prinzipiell jede DNA als Template geeignet ist, an die die kurzen SfiI-Schnittstellen heranamplifiziert werden können.

3.3.3.2 Antikörpersequenzierung

Zur Diversitätsbestimmung von Antikörperrepertoires verschiedener Spender sollte für die Arbeitsgruppe Konthur (MPI für molekulare Genetik, Berlin) ein Deepsequencing auf dem 454 durchgeführt werden. Da aufgrund der gewählten Herangehensweise und der damit verbundenen Primerwahl eine targetspezifische und gerichtete Libraryerstellung und Sequenzierung benötigt wurden (Prinzip der Methode siehe Kapitel 2.2.7.10 auf Seite 59 f), konnten die 454-Librarys nur nach Standard-Amplicon-Protokoll oder nach der neuen Methodik hergestellt werden.

Versuche zur Erstellung einer Standard-Amplicon Library scheiterten (Ergebnisse werden nicht gezeigt; Versuchsdurchführung Florian Rubelt, Arbeitsgruppe Konthur, MPI für molekulare Genetik, Berlin). Also wurde die neue Methodik zur Libraryherstellung auch für die Sequenzierung von Immunglobulin-cDNAs herangezogen; es sollte überprüft werden, ob sie auch darauf anwendbar ist und brauchbare 454-Runs liefert. Im folgenden Kapitel werden die exemplarischen Ergebnisse einer Sequenzierung von vier Librarys auf einer PTP-Region demonstriert.

Libraryherstellung. Aus der gesamt-RNA peripherer Blutzellen wurde von Florian Rubelt (AG Konthur, MPI für molekulare Genetik, Berlin) die Erststrangsynthese sowie ein Heranampifizieren der SfiI-Sites durchgeführt. Darauf erfolgte die 454-Library-Herstellung nach der neuen Methodik. Abb. 86 zeigt die Ergebnisse einer Library auf einem Agilent Bioanalyzer 2100 High Sensitivity DNA Chip.

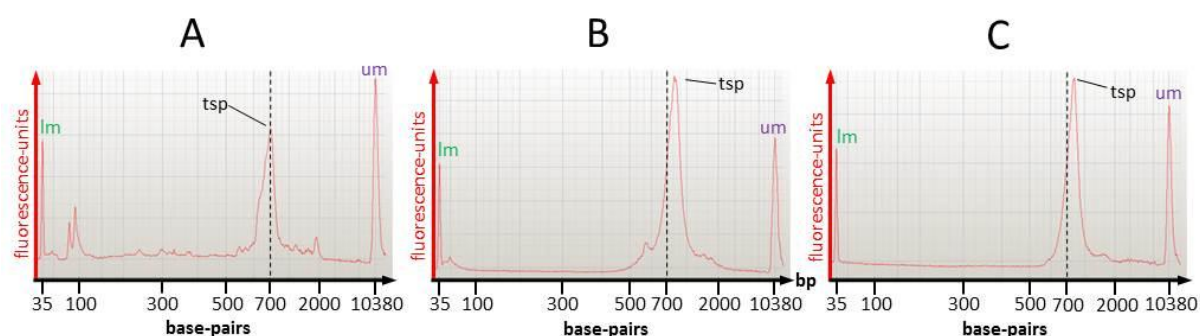


Abb. 86: Größenverteilung der Antikörper-Librarys

Dargestellt sind die Konzentrationsprofile einer IgH-cDNA-Library zu bestimmten Zeitpunkten der Libraryproduktion. Die IgH-cDNA ist als Target-spezifisches-Produkt (tsp) gekennzeichnet. (A) zeigt das Konzentrationsprofil nach durchgeführter PCR zur targetspezifischen Amplifikation und Addition der SfiI-Sites. In (B) ist das Profil der Library nach Gelausschnitt und erster Ampurebead-Aufreinigung zu sehen und (C) zeigt das Konzentrationsprofil nach zweiter Ampurebead-Aufreinigung (Abb. aus Knaust et al, submitted).

Anhand des Konzentrationsprofils der Library nach der PCR (Abb. 86A) lässt sich erkennen, dass neben dem spezifischen Produkt (tsp) mit einer erwarteten Länge um die 700 bp andere Produkte vorhanden sind. Es sind knapp oberhalb von 35 bp und besonders um 100 bp deutliche Peaks sichtbar; auch im Bereich von 100 bp bis hoch zum tsp sind immer wieder andere Amplifikationsprodukte zu erkennen.

Abb. 86B zeigt das Profil zum Zeitpunkt nach erfolgtem Gelausschnitt, Restriktion und Ligation einschließlich ihrer beiden MinElute-Aufreinigungen und einer ersten Ampurebead-Aufreinigung. Der Peak des tsp ist länger als zuvor (Erweiterung um die 454-Sequenzbereiche) und es fehlen die

deutlichen Peaks bei 100 bp und länger. Lediglich die Peaks knapp oberhalb 35 bp und zwischen 500 bp und dem tsp sind sichtbar.

Nach der zweiten Ampurebead-Aufreinigung (Abb. 86C) ist auf dem Konzentrationsprofil nur noch der tsp erkennbar. Die Library konnte erfolgreich aufgereinigt werden und ist zur Sequenzierung geeignet.

emPCR, Beadrecovery und Sequencing. Vier Librarys wurden in einer emPCR in einem Cup zu gleichen Anteilen angesetzt und im Beadrecovery aufgereinigt. Alle erhaltenen Beads (1,95 Mio, entspricht einer Enrichment-Rate von 5%) wurden auf eine PTP-Region aufgetragen und sequenziert. Tab. 43 zeigt die Ergebnisse der Region, in den folgenden Abbildungen sind sie graphisch dargestellt; Tab. 44 und Abb. 90 zeigen eine Analyse der Kontroll-DNA. Tab. 45 beschreibt die Zerlegung der erzeugten Reads nach ihren MIDs und in Abb. 91 ist die Größenverteilung der Librarys dargestellt.

Bei diesem Run erzeugte Region 1 über 410000 Reads mit einem Anteil von 43% (Passed Filter Wells) an den Key-Pass-Wells bei insgesamt knapp über 1 Mio. Raw Wells (Tab. 43). Die durchschnittliche Leseweite der Library-Reads liegt bei gut 378 bp mit entsprechender Verteilung (Abb. 87). Der Anteil der Failed Short Quality Reads von 23% ist minimal erhöht, die Failed Dot und Mixed sind mit 34% schon etwas größer als bei einem idealen Run (siehe Tab. 43 und Abb. 89 auf Seite 114), wobei die Fraktion der Failed-Dots mit knapp 200000 Reads den größeren Anteil stellt.

Die Ergebnisse der Kontroll-DNA-Reads liegen innerhalb der Vorgaben von Roche; ihre durchschnittlichen Leseweiten (Tab. 43) sind gut bei entsprechender Größenverteilung (Abb. 88 auf Seite 113). Die Filterwerte mit über 95% Passed Filter Wells für beide Chargen sind sehr gut (Tab. 43 und Abb. 89 auf Seite 114). Die Anteile über alle Längen bei den verschiedenen Übereinstimmungen sind sehr hoch, die Qualität der Kontroll-DNA-Reads über ihre gesamte Länge ist sehr gut (Tab. 44 und Abb. 90 auf Seite 114).

Die Zerlegung der Reads auf Region 1 nach ihren MIDs (Tab. 45 auf Seite 114) zeigt eine relativ gleichmäßige Verteilung der vier Antikörperlibrarys. Ihre Anteile liegen zwischen 17% und 28%, was dem berechneten Ansatz in der emPCR (gleiche Anteile, also 25%) nahe kommt. Die verschiedenen Größenverteilung der vier Librarys (Abb. 91 auf Seite 115) spiegelt die Größenverteilung aller Reads während des Runs wider; keine fällt negativ durch einen zu hohen Anteil an kurzen Reads auf. Bei allen Librarys wird die größte Fraktion von Reads zwischen 400-500 bp gebildet und der Anteil an Reads größer als 400 bp liegt bei mindestens 46%. Der Anteil an kurzen Reads (bis 100 bp) ist nicht größer als 12%, und bei einer IgH-Library (hpC) liegt er sogar nur bei 4%.

Der Großteil der Reads mit ca. 100 bp Länge trug als Insert nur eine Base und bestand ansonsten aus den Primersequenzen und 454-Adapterbereichen (Analyse zur Verfügung gestellt von Florian Rubelt, AG Konthur, Max-Planck-Institut für molekulare Genetik, Berlin).

	GACT Region 1	CATG-Wells Region 1	ATGC-Wells Region 1
Raw Wells	1018280	1018280	1018280
Key Pass Wells	956547	10920	14983
Passed Filter Wells	410627	10405	14607
Total Bases	155179932	4079595	7127549
% Passed Filter	42,93	95,28	97,49
Lenght Average	377,91	392,08	487,95
Lenght Std Deviation	140	87,47	67,65
Longest Reads Lenght	709	544	654
Shortest Reads Lenght	40	40	41
Median Reads Lenght	429	409	512
Failed Dot	195689	154	174
Failed Mixed	126795	8	32
Failed Short Quality	222176	299	342
Failed Short Primer	1087	0	0
% Dot + Mixed	33,71	1,48	1,37
% Short	23,34	2,74	2,28

Tab. 43: Run der Antikörpersequenzierung

Es sind die Ergebnisse von Region 1 dargestellt, die von der GSRUNBrowser-Software ausgegeben werden. Die GACT-Spalte zeigt die Library-Reads, die Spalten der CATG- und ATGC-Wells die der Kontroll-DNA-Reads.

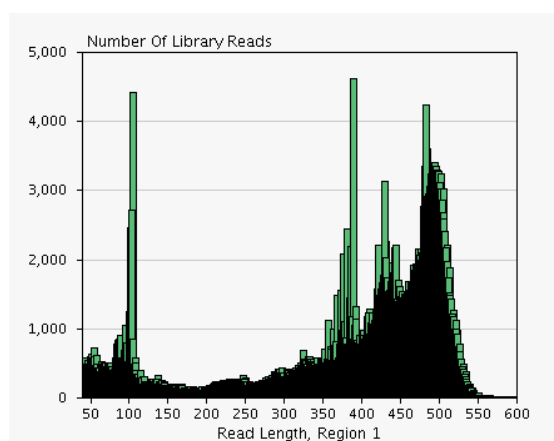


Abb. 87: Größenverteilung der Antikörper-Library-Reads

Ausgegeben von der GSRUNBrowser-Software.

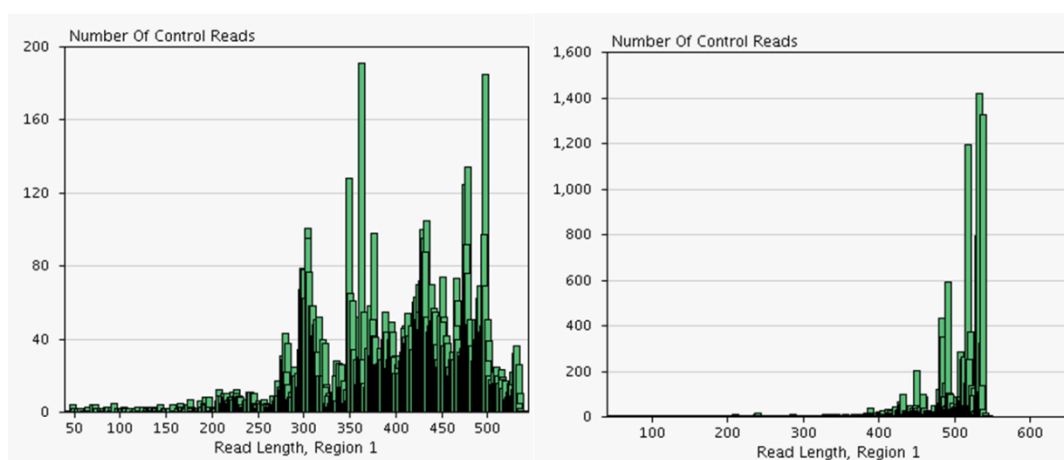


Abb. 88: Größenverteilungen der Kontroll-DNA-Reads

Es werden die von der GSRUNBrowser-Software ausgegebenen Verteilungen der CATG- (links) und ATGC- Kontroll-DNA-Reads (rechts) dargestellt.

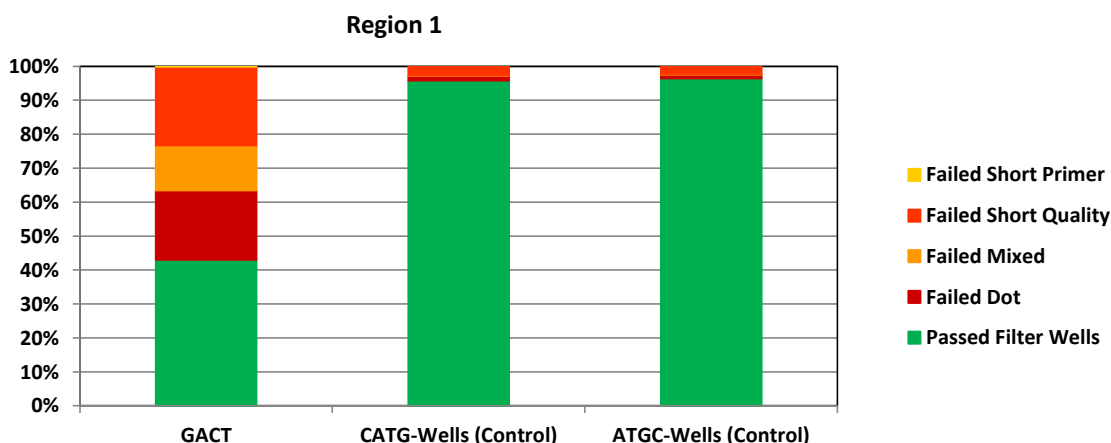


Abb. 89: Prozentuale Verteilung beim Run mit den Antikörper-Librarys

Das Verhältnis von Passed Filter Wells (grün) zu den Failed Wells wird für Region 1 gezeigt, für die Library-Reads (gact-Wells) und die Kontroll-DNA-Reads (catg und atgc-Wells).

Länge	100 bp			200 bp			300 bp			400 bp		
Match	100%	98%	95%	100%	98%	95%	100%	98%	95%	100%	98%	95%
Region 1	83,97	94,56	96,32	78,69	94,41	95,95	47,30	90,71	94,31	26,14	81,33	90,37

Tab. 44: Kontroll-DNA-Analyse des Runs mit den Antikörper-Librarys

Es werden die von der GSRUNBrowser-Software ausgegebenen Ergebnisse der Kontroll-DNA-Analyse dargestellt. Für beide Regionen des Runs wird der prozentuale Anteil der Kontroll-Bead-Reads gezeigt, der zu je 100%, 98% und 95% mit den Kontroll-DNA-Sequenzen über eine Länge von 100 bp, 200 bp, 300 bp und 400 bp übereinstimmt (Match).

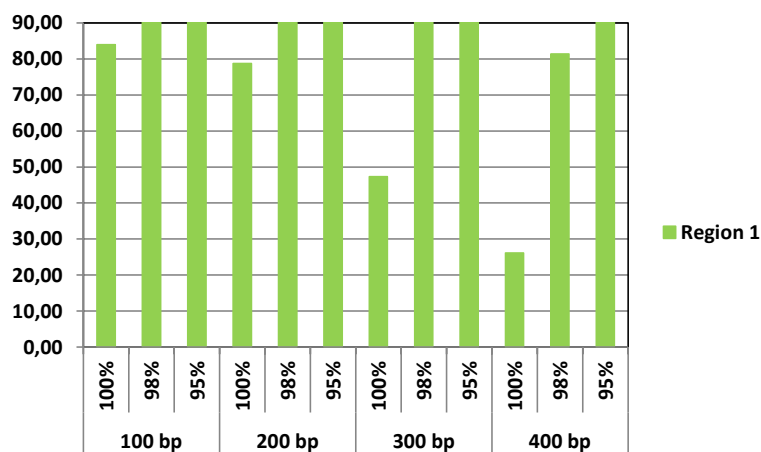


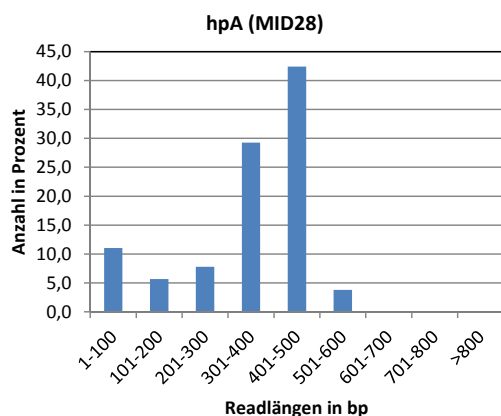
Abb. 90: Graphische Darstellung der Kontroll-DNA-Analyse

Es werden die Ergebnisse aus Tab. 44 graphisch dargestellt.

MID	28	32	35	39	non-matching
Library	hpA	nd32H	hpB	hpC	-
Read-Anzahl	115022	107515	90971	72313	24806
Anteil in Prozent	28,01	26,18	22,15	17,61	6,04

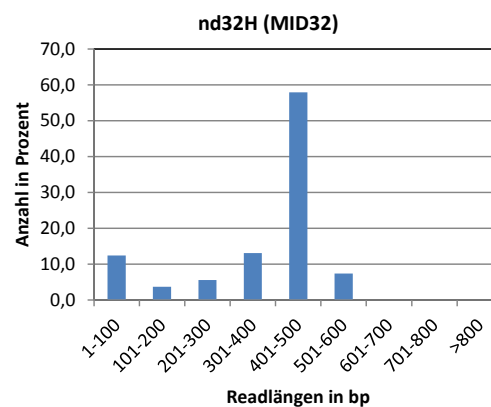
Tab. 45: Zerlegung in die MIDs beim Run mit den Antikörper-Librarys

Angegeben ist das Ergebnis für Region 1, auf der die Antikörper-Librarys geladen waren. Das verwendete Mismatch beim MID-Splitting ist Null.



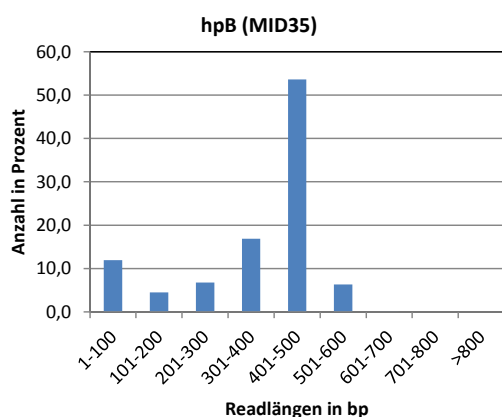
Gesamte Read-Anzahl: 115022
Durchschnittl. Read-Länge: 352 bp

Readlängen in bp	Anzahl der Reads	Anteil in Prozent
1-100	12703	11,0
101-200	6524	5,7
201-300	8965	7,8
301-400	33653	29,3
401-500	48808	42,4
501-600	4369	3,8
601-700	0	0,0
701-800	0	0,0
>800	0	0,0



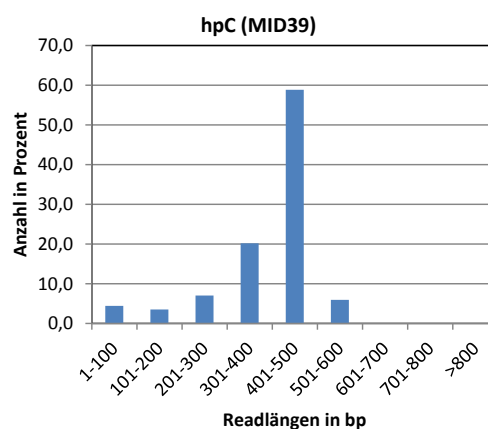
Gesamte Read-Anzahl: 107515
Durchschnittl. Read-Länge: 379 bp

Readlängen in bp	Anzahl der Reads	Anteil in Prozent
1-100	13376	12,4
101-200	3932	3,7
201-300	5957	5,5
301-400	14044	13,1
401-500	62259	57,9
501-600	7947	7,4
601-700	0	0,0
701-800	0	0,0
>800	0	0,0



Gesamte Read-Anzahl: 90971
Durchschnittl. Read-Länge: 371 bp

Readlängen in bp	Anzahl der Reads	Anteil in Prozent
1-100	10830	11,9
101-200	4087	4,5
201-300	6130	6,7
301-400	15358	16,9
401-500	48784	53,6
501-600	5751	6,3
601-700	1	0,0
701-800	0	0,0
>800	0	0,0



Gesamte Read-Anzahl: 72313
Durchschnittl. Read-Länge: 398 bp

Readlängen in bp	Anzahl der Reads	Anteil in Prozent
1-100	3214	4,4
101-200	2551	3,5
201-300	5104	7,1
301-400	14604	20,2
401-500	42549	58,8
501-600	4289	5,9
601-700	2	0,0
701-800	0	0,0
>800	0	0,0

Abb. 91: Größenverteilungen der Antikörper-Librarys im Run

Bei der Angabe der durchschnittlichen Read-Länge wurde das 10 bp lange MID abgezogen.

Aufgrund der erhaltenen Reads und der durchschnittlichen Leseweiten der Library-Reads kann auch dieser Run als Erfolg gewertet werden. Die Antikörper-Librarys lieferten gute Sequenzierungsergebnisse; die neue Methodik zur 454-Library-Erstellung kann also auch zur Immunglobulinsequenzierung aus Gesamt-RNA erfolgreich angewendet werden und stellt eine Alternative zum Standard-Amplicon-Protokoll dar, das für diesen Ansatz zu keinem Ergebnis führte.

Mit dieser Methode konnten in mehreren 454-Runs rund 3,57 Mio. Reads aus IgH- und auch IgLK-Librarys generiert werden, was in einer Publikation mündete [Rubelt et al. accepted 2012].

3.3.3.3 Verbesserungen des Protokolls für die Antikörpersequenzierung

Im Zuge der Zusammenarbeit mit der AG Konthur bei der Antikörpersequenzierung stellte sich heraus, dass die Templatemengen, aus denen 454-Librarys zum Immunglobulin-Deepsequencing erstellt werden sollten, teilweise sehr gering ausfielen; oft zu gering, um eine erfolgreiche Library-Produktion nach dem bisherigen Protokoll zu starten. Engpässe stellten die geringen Gesamt-RNA-Mengen oder die Ausbeuten bei der PCR dar (Rücksprache mit Florian Rubelt). Das bisherige Protokoll musste verändert werden, um die 454-Library-Erstellung aus wesentlich geringeren Templatemengen (kleiner als 200 ng) zu erlauben (vergleiche Seite 62 ff.).

Libraryerstellung. Es wurden aus cDNAs mittels des veränderten Protokolls 454-Librarys hergestellt und die Konzentration vor und nach der 454-Library-Präparation gemessen (Tab. 46).

Library Name	Menge in ng zu Beginn	tMID	Menge in ng am Ende	Konzentra- tion in ng/μl
RA-BS_03LK	6,66	29	0,15	0,02563
RA-BS_10LK	9,13	31	0,10	0,01611
RA-BS_02H	11,85	32	0,69	0,11441
RA-BS_07H	39,96	26	8,33	1,38817
RA-BS_14LK	23,32	33	5,19	0,86453
Kareti LK	29,71	23	2,34	0,38929
RA-BS_12H	59,26	27	15,42	2,57035
TM51H	38,6	24	2,50	0,416
ND07LK	20	25	0,59	0,09901

Tab. 46: Mengenvergleich vor und nach Libraryproduktion

Die Mengen zu Beginn (vor der Restriktion) wurden fluorometrisch mit dem Qubit bestimmt, die Mengen am Ende errechnen sich aus den High-Sensitive-DNA-Chip-Messungen entnommenen Konzentrationen und dem Volumen der Library (6 μl nach Einengen).

Vor dem SfiI-Restriktionsverdau lagen die cDNAs in Mengen von 59 ng bis zu nur 6,6 ng vor; am Ende der 454-Library-Produktion waren sie noch in Mengen zwischen 15 ng und 100 pg vorhanden.

Die mittels High-Sensitiv-Chip gemessenen Konzentrationsprofile der Librarys RA-BS_10LK (Abb. 92), KaretiLK (Abb. 93) und KaretiLK nach dem optionalen Größenunterschuss (Abb. 94) werden als Beispiele gezeigt.

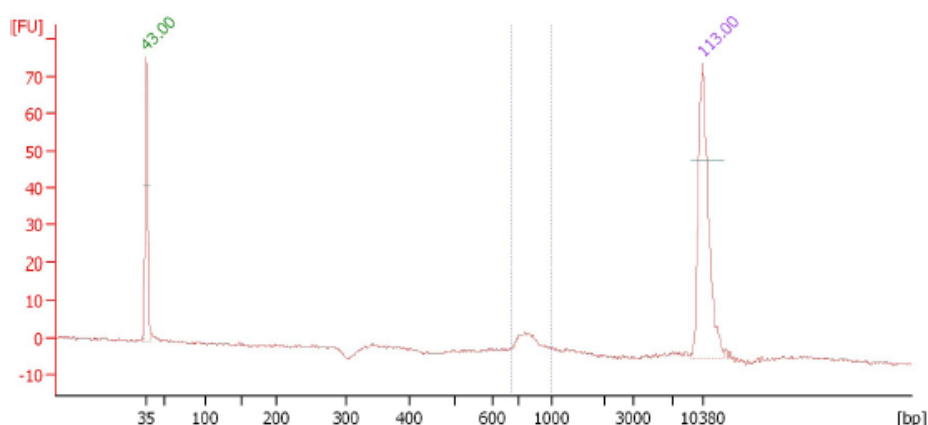


Abb. 92: Konzentrationsprofil der Library RA-BS_10LK

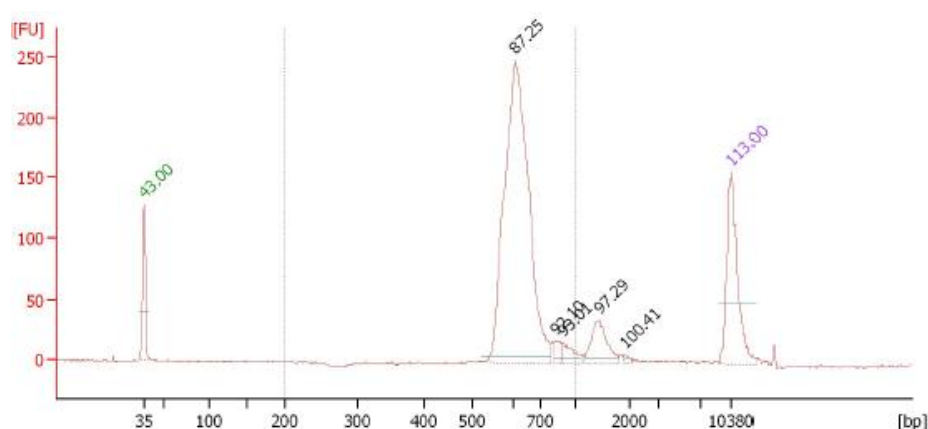


Abb. 93: Konzentrationsprofil der Library KaretikLK

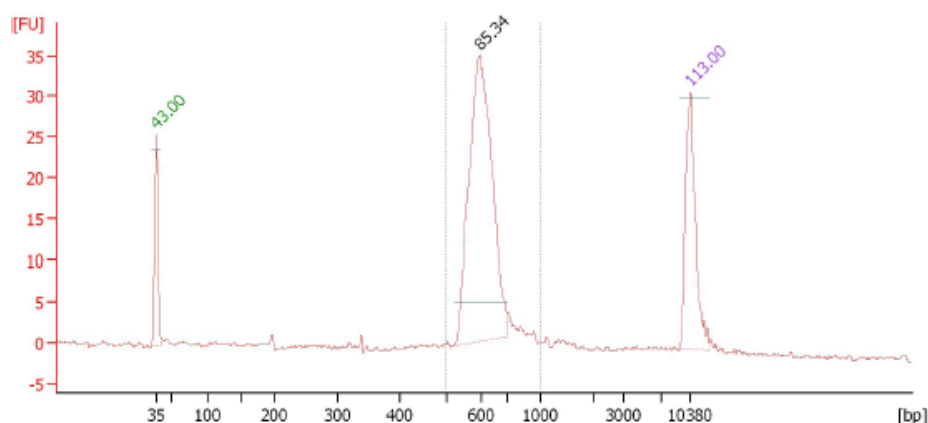


Abb. 94: Konzentrationsprofil der Library KaretikLK nach optionalem Größenunterschluss

Bei Library RA-BS_10LK ist im Bereich von 700 bp, was die erwartete Größe des target spezifischen Produktes darstellt, nur ein schwacher Peak zu sehen; diese Library hatte die geringste DNA-Menge zum Ende der Library-Herstellung (Tab. 46). Die KaretikLK-Library (Abb. 93) zeigte zum Ende der Library-Produktion nach der doppelten SPRI-Bead-Aufreinigung einen deutlichen Peak bei der erwarteten Länge von ca. 600 bp des spezifischen Amplifikationsproduktes; allerdings ist auch ein weiterer deutlicher Peak knapp unterhalb von 2000 bp zu sehen. Daher wurde mit dieser Library ein optionaler Größenunterschluss durchgeführt, um die Fragmente knapp unterhalb 2000 bp zu

entfernen. Nach dem Größenunterschuss (Abb. 94 auf Seite 117) ist nur noch der Peak des spezifischen Amplifikationsproduktes zu sehen.

emPCR, Beadrecovery und Sequencing. In einem emPCR-Cup wurden die nach verbessertem Protokoll hergestellten Librarys angesetzt (siehe Abb. 40 auf Seite 63) und alle (1,7 Millionen) im Beadrecovery erhaltenen Beads auf die Region einer PTP geladen (die Beads hatten eine Enrichment-Rate von 6,5%). Die Ergebnisse des Runs sind in Tab. 47 und den folgenden Abbildungen dargestellt; Tab. 48 und Abb. 98 zeigen die Kontroll-DNA-Ergebnisse.

	GACT Region 1	CATG-Wells Region 1	ATGC-Wells Region 1
Raw Wells	984621	984621	984621
Key Pass Wells	936469	7091	7573
Passed Filter Wells	410667	6602	7315
Total Bases	156179308	2797054	3575273
% Passed Filter	43,85	93,10	96,59
Lenght Average	380,31	423,67	488,75
Lenght Std Deviation	157,55	104,65	83,27
Longest Reads Lenght	1196	590	561
Shortest Reads Lenght	40	46	46
Median Reads Lenght	467	458	518
Failed Dot	139554	137	115
Failed Mixed	150052	7	14
Failed Short Quality	230072	293	109
Failed Short Primer	20	0	0
% Dot + Mixed	30,93	2,03	1,70
% Short	24,57	4,13	1,44

Tab. 47: Run mit Librarys nach verbessertem Protokoll

Dargestellt sind die von der GSRUNBrowser-Software ausgegebenen Ergebnisse von Region 1; die linke Spalte (GACT-Wells) zeigt die Library-Reads. Die mittlere und rechte Spalte stellen die Kontroll-DNA-Reads dar –es fand ein Wechsel zu zwei verschiedenen Kontroll-DNA-Bead-Chargen statt, weshalb zwei Kontroll-DNA-Keys aufgelistet sind (catg und atgc).

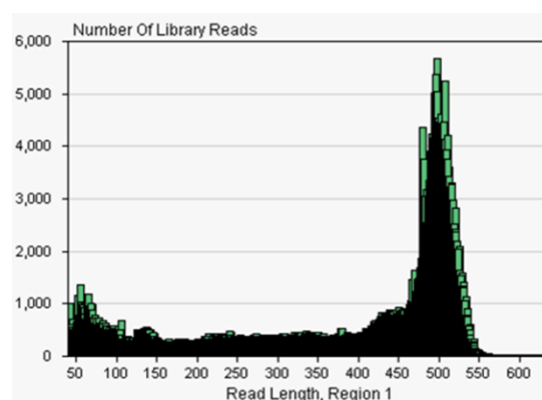


Abb. 95: Größenverteilung der Libraryreads

Ausgegeben von der GSRUNBrowser-Software.

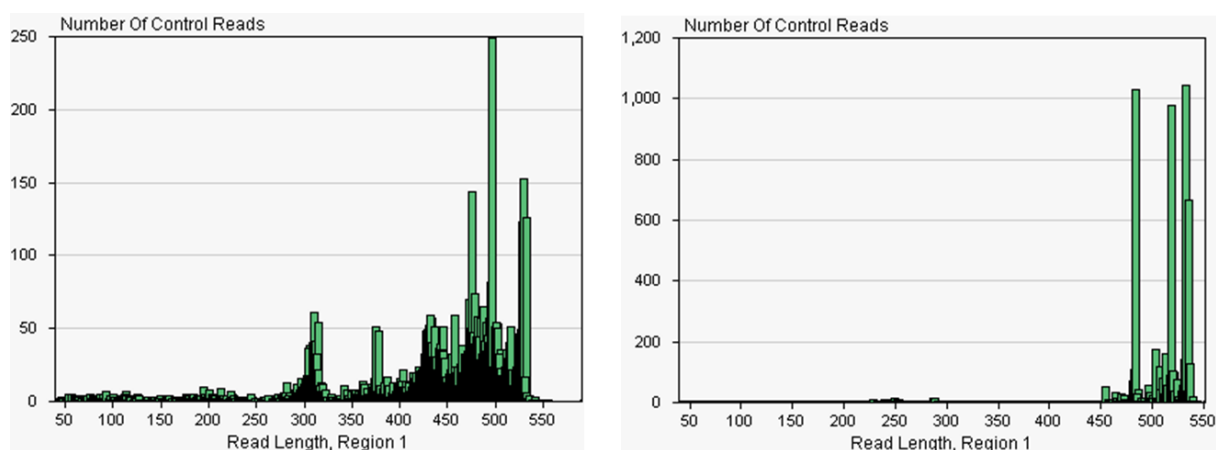


Abb. 96: Größenverteilung der Kontroll-DNA-Reads

Die von der GSRUNBrowser-Software ausgegebenen Verteilungen für die catg-Wells (links) und atgc-Wells (rechts).

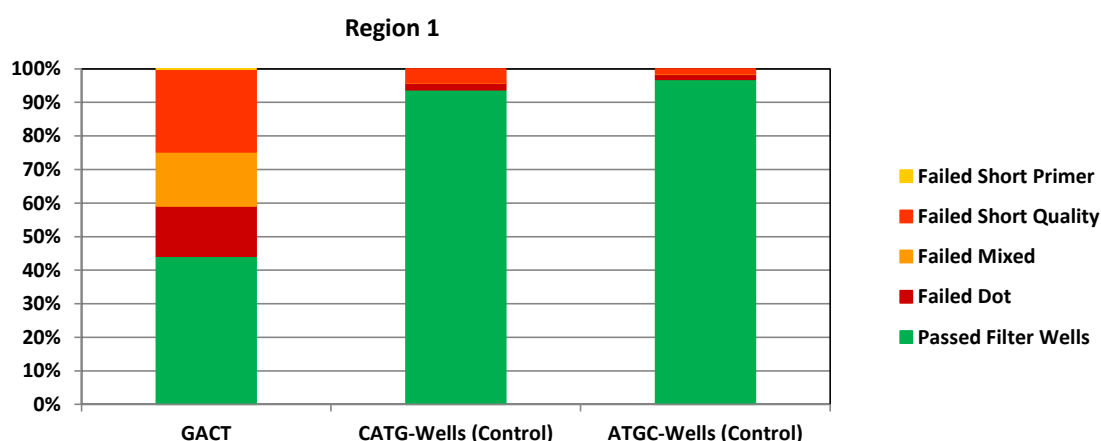


Abb. 97: Prozentuale Verteilung beim Run

Für Region 1 wird das Verhältnis von Passed Filter Wells (grün) zu den Failed Wells dargestellt; für die Library-Reads (gact-Wells) und die Kontroll-DNA-Reads (catg und atgc-Wells).

Länge	100 bp			200 bp			300 bp			400 bp		
Match	100%	98%	95%	100%	98%	95%	100%	98%	95%	100%	98%	95%
Region 1	80,03	88,69	90,95	73,29	86,95	88,71	53,93	84,68	87,83	37,61	78,23	85,07

Tab. 48: Kontroll-DNA-Analyse

Für Region 1 sind die von der GSRUNBrowser-Software ausgegebenen Ergebnisse der Kontroll-DNA-Analyse angegeben. Angezeigt werden die prozentualen Anteile der Kontroll-Bead-Reads, die zu je 100%, 98% und 95% mit den Kontroll-DNA-Sequenzen über Längen von 100 bp, 200 bp, 300 bp und 400 bp übereinstimmen (Match).

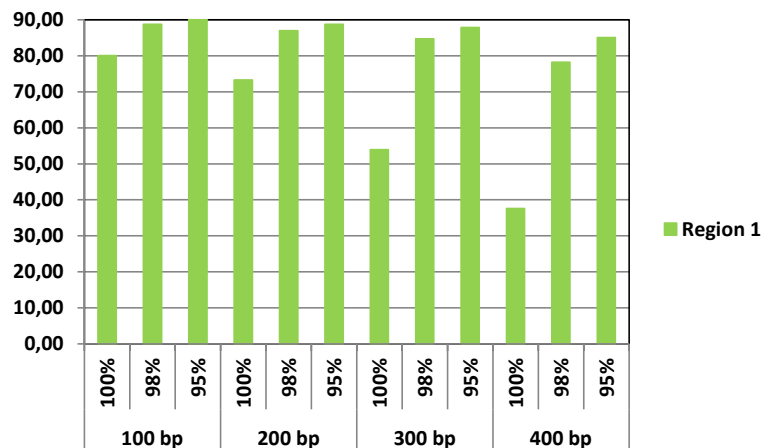


Abb. 98: Graphische Darstellung der Kontroll-DNA-Analyse

Region 1 erzeugte bei diesem Run über 410000 Reads mit einem Anteil von rund 44% (Passed Filter Wells) an den Key-Pass-Wells; insgesamt wurden 984621 Raw Wells gefunden (siehe Tab. 47 auf Seite 118). Die durchschnittliche Leseweite liegt bei 380 bp, der Median bei 467 bp, was sich in einer entsprechend rechtslastigen Verteilung widerspiegelt (Abb. 95 auf Seite 118). Bei den Failed Reads sind die Anteile von sowohl Dot- und Mixed-, als auch der „Failed Short Quality“-Reads leicht erhöht, wobei letztere die größte Fraktion stellen (vergleiche siehe Tab. 47 auf Seite 118 und Abb. 97 auf Seite 119).

Die Kontroll-DNA-Reads liegen innerhalb der Vorgaben von Roche, sie zeigen gute durchschnittliche Leselängen (Tab. 47) und Verteilungen (Abb. 96 auf Seite 119) mit sehr guten Filterwerten von über 93% Passed Filter Wells für beide Kontroll-DNA-Chargen (Tab. 47 auf Seite 118 und Abb. 97 auf Seite 119). Auch der Anteil bei den verschiedenen Übereinstimmungen über alle Längen ist hoch, selbst bei 400 bp haben immer noch knapp 78% aller Kontroll-DNA-Reads eine 98%ige Übereinstimmung zu der erwarteten Sequenz (Tab. 48 auf Seite 119); die Qualität der Kontroll-DNA-Reads ist über die gesamte Länge gut (Abb. 98).

Die erhaltenen Readmengen mit der guten durchschnittlichen Leseweite sprechen für einen erfolgreichen Run und zeigen, dass die Verbesserungen des Protokolls für die Verringerung der Templatemengen funktionieren.

4. Diskussion

4.1 Ampurebead-Titration

Um gute Run-Ergebnisse beim 454-Sequencing mit dem Titanium-System zu erhalten und die anfallenden Kosten zu minimieren, muss die durchschnittliche Leselänge der auf dem GS-Sequencer generierten Reads groß sein; in Verbindung mit einer möglichst großen Anzahl an generierten Reads ergibt das eine hohe Ausbeute und entsprechend große Datenmenge. Eine Grundvoraussetzung dafür ist, dass die zu sequenzierenden Librarys in dem von Roche als ideal bezeichneten Größenbereich von 500-900 bp vorliegen. Größere DNA-Fragmente lassen sich nur schlecht in der emPCR amplifizieren, führen dadurch zu Beads, die aufgrund einer geringeren Anzahl von DNA-Amplifikationsprodukten an den Capture-Beads entweder während des Enrichment beim Beadrecovery entfernt werden, oder zu schwache Signalintensitäten im Sequencing liefern. Fragmente kleiner 500 bp stellen ein doppeltes Problem dar: ihre DNA-Insert-Sequenz, also das, was sequenziert wird, ist kurz und kann so die Datenmengen-Ausbeute auf dem 454-Sequencer empfindlich verringern; außerdem lassen sich die kürzeren Fragmente in der emPCR besser amplifizieren und führen zu Fehlkalkulationen bei Library-Konzentrations-Berechnungen, was wiederum zu hohe Enrichment-Rates und damit meistens fürs Sequencing unbrauchbare DNA-Beads bedeutet. Ferner können die Beads, die kürzere Fragmente tragen, im Sequencing zu einer Verschlechterung der Readlängen anderer Beads mit längeren Fragmenten führen, da deren Signalintensität geringer sein könnte und von den stärkeren Signalen anderer Beads auf der PTP überstrahlt werden könnten.

Um 454-Librarys von kürzeren Fragmenten zu befreien, werden diese mit SPRI-Beads aufgereinigt. Die Vorteile dieser Methode liegen darin, dass sie schnell durchführbar ist, im Vergleich zu gängigen Säulen mit Ionen-Austauschern geringere Ausbeute-Verluste besitzt und die erwünschten DNA-Größen in einem bestimmten Bereich frei gewählt werden können. Letzteres ist durch die Funktionsweise der SPRI-Beads begründet; die Konzentration an „Crowding-Reagents“ ist dafür entscheidend, welche Größen an die SPRI-Beads getrieben werden und welche in Lösung verbleiben. Damit man diese Größenselektion steuern kann, wurden die verwendeten SPRI-Bead-Chargen (Ampurebeads) titriert.

Von Roche wird die Konzentration des Marker-Peaks bei 900 bp als Bezugsgröße verwendet, mit der die Konzentrationen der kleineren Peaks ins Verhältnis gesetzt werden. Eine solche Bezugsgröße ist notwendig, da es bei der Versuchsdurchführung zu manuellen Fehlern, hauptsächlich verursacht durch kleine Abweichungen beim Pipettieren der unterschiedlichen Ansätze, kommen kann, weshalb keine absoluten Messergebnisse verwendet werden. Der als Bezugsgröße verwendete Peak muss die Eigenschaften besitzen, dass er bei den unterschiedlichen Ampurebead-Aufreinigungen nicht größenselektiv entfernt wird, egal ob kleine Volumina SPRI-Beads (50 µl) oder größere verwendet werden -er sollte idealerweise von der Ampurebead-Aufreinigung unangetastet bleiben; außerdem

sollte er eindeutig auf dem 7500-DNA-Chip identifizierbar sein. Beides trifft auf den Peak bei 900 bp zu.

Die Vorgaben von Roche für die Peak-Ratios zur Library-Aufreinigung müssen einen Kompromiss darstellen: kleine, nicht gewollte Fragmente müssen so gut es geht ausgeschlossen werden, und die größeren Fragmente (ab 500 bp) sollten ohne große Verluste die Aufreinigung durchlaufen. Wählt man die Bedingungen für die Ampurebead-Aufreinigung zu strikt, erhält man zwar einen guten Ausschluss kleinerer Fragmente, wie bei den Volumen-Verhältnisse von 50:100 oder 55:100 geschehen -hier liegen messbar keine Fragmente kleiner 400 bzw. 500 bp vor; allerdings verliert man auch in erheblichem Maß größere Fragmente (man vergleiche die entsprechenden Peak-Ratios von 500/900 mit denen der Kontrolle –bei dem Verhältnis von 50:100 übersteht von den 500 bp-Fragmenten nur rund ein Fünftel die Aufreinigung). Vergleicht man die gemessenen Peak-Ratios der unterschiedlichen Volumenverhältnisse mit den Vorgaben von Roche, so liegen die Roche-Vorgaben bei dieser Ampurebead-Titration zwischen den Volumenverhältnis von 60:100 und 65:100, also zwischen 60 und 65 µl Ampurebeads pro 100 µl DNA.

Aus vorangegangenen Sequencing-Projekten mit dem 454 [124] und der dafür erforderlichen Herstellung von genomischen 454-Librarys, bei der SPRI zur Größenselektion verwendet wurde, wurde die Erfahrung gemacht, dass es besser ist, etwas unterhalb der Vorgaben von Roche zu bleiben: wie aus der Messung auf dem 7500 DNA-Chip ersichtlich ist, reichen wenige Mikroliter Unterschied einer Komponente, um durch die Änderung des SPRI-Bead-DNA-Verhältnisses eine Verschiebung in der Fragment-Aufreinigung zu erreichen. Zu solchen Unterschieden bzw. Fehlern von wenigen Mikroliter beim Pipettieren kann es im Laboralltag durchaus kommen, da die Ampurebead-Suspension viskos ist und beim Pipettieren außen an der Pipettenspitze immer etwas Suspension mitkommt, was eine Vergrößerung des Ampurebead-Volumens und damit eine Verschiebung zu dem Verbleib kleinerer Fragmente darstellt. Um das Entfernen kleinerer Fragmente sicherzustellen, wurde bei dieser Charge Ampurebeads aufgrund dieser Erfahrungen und den Ergebnissen der Titration auf dem 7500 DNA-Chip eine Menge von 61 µl Ampurebeads auf 100 µl DNA bei den SPRI-Aufreinigungen als ideal für den Größenausschluss von Fragmenten kleiner 500 bp angesehen und verwendet.

4.2 Sequenzierungsversuche des aqf-Projektes

Das aqf-Projekt sah vor, Tausende von cDNA-Sequenzen mit NGS-Technologie zu erzeugen, um einen Vergleich mit Standard Microarrays für das Auffinden von ESTs zu ermöglichen. Die Sequenzierung der aqf-cDNAs war das erste Projekt im Max Planck Institut für molekulare Genetik, bei dem cDNAs auf dem 454 mit der Titanium-Chemie sequenziert werden sollten. Zu dem Zeitpunkt gab es von Roche keine spezielle Methode für das Sequenzieren von cDNAs bzw. für das Erstellen von 454-Librarys aus cDNAs. Zu dem Zeitpunkt existierten nur das Titanium General Library-Protokoll und ein Technical Bulletin zur Herstellung von Amplicon Librarys. Ein Sequenzieren mit dem älteren FLX-Standardsystem erübrigte sich, da es auch dort nur diese entsprechenden Protokolle gab.

Auch zum heutigen Zeitpunkt gibt es von Roche nur zwei Protokolle für das Sequenzieren von cDNAs: ebenfalls ein Amplicon-Protokoll, was praktisch dem entspricht, was es zu dem Zeitpunkt des aqf-Projektes gab -wobei das neue lediglich um 5 bp kürzere Adaptersequenzen verwendet, ansonsten ist es identisch; und ein Rapid-Library-Protokoll, bei dem die RNA fragmentiert und dann über Random-Primer in cDNA überschrieben wird; anschließend wird aus diesen Fragmentierungs-Produkten eine Rapid-Library erstellt. An Fragmente wird mittels Random-Ligation von 454-Adapter eine Library erstellt –das ist dasselbe Prinzip wie bei der Erstellung einer Titanium General Library (nur das letzteres länger dauert, mehr Schritte enthält und größere Ausgangsmengen braucht).

Damals gab es also nur die Möglichkeit, entweder aus den cDNAs eine Amplicon Library zu erstellen oder die cDNAs wie genomische DNA zu behandeln, also eine General Library zu erstellen für die Sequenzierung. Da es zu diesem Zeitpunkt keine Erfahrungen mit beiden Wegen für die cDNA-Sequenzierung gab –die Titanium-Chemie war gerade eingeführt worden, wurden beide Wege ausprobiert und verglichen. Die Arbeitsgruppe Kube stellte die cDNA mit dem MINT-Kit her; die cDNAs besitzen die durch das MINT-Kit eingeführten flankierenden Sequenzen und Längen von 500 bp bis 2000 bp. Da die Erststrangsynthese beim MINT-Kit über das Poly-A der RNAs durch einen (T)₂₀-Primer stattfindet, besitzen die cDNAs im Bereich des 3'-Endes eine Homopolymere Region von Adenin (bzw. von Thymin im 5'-Bereich des Gegenstranges).

4.2.1 Sequenzierung von aqf-cDNAs über die Erstellung einer General Library

Zuerst wurde versucht, die aqf-cDNAs über die Erstellung einer Titanium General Library zu sequenzieren.

Für einen Größenbereich der cDNA von 500 bp bis 2000 bp gibt es prinzipiell zwei Möglichkeiten bei der Erstellung einer General-Library: man fragmentiert die cDNA oder nicht. Man könnte das Fragmentieren auslassen und riskieren, dass die größeren Fragmente (größer als 1000 bp) kaum oder schlecht sequenziert werden; der Vorteil läge in einer geringeren Input-DNA-Menge und der Möglichkeit, die cDNAs in voller Länge in einem Stück zu sequenzieren (ohne Assembly). Es gibt aber einen entscheidenden Nachteil dabei, weshalb eine Fragmentierung durchgeführt wurde: die cDNAs besitzen einen Poly-A-Bereich (bzw. Poly-T auf dem Gegenstrang). Da bei der Erstellung einer General Library die 454-Adapter über Random-Ligation an die Insert-DNA gebracht wird, würden ohne Fragmentierung 50% der sequenzierbaren Inserts am Anfang einen Poly-T-Bereich besitzen. Pyro-Sequencing ist aber nicht geeignet, um über solche langen homopolymeren Bereiche zu lesen; das Mitführen solcher Sequenzbereiche würde zu schlechten Ergebnissen im Sequencing führen. Durch Fragmentierung sollte dieser prozentuale Anteil gesenkt werden können (als Beispiel: eine 2 kb lange cDNA wird an einer mittleren Stelle während der Fragmentierung gebrochen, es entstehen daraus zwei Fragmente; der Anteil an sequenzierbaren Inserts, die am 5'-Anfang eine Poly-T-Region besitzen, liegt dann bei 25%).

4.2.1.1 Erstellung der Titanium General Library

Zur Fragmentierung der vier cDNAs aqf3b, aqf4b, aqf5b und aqf6b wurde Ultraschall angewendet und nicht wie im Protokoll von Roche vorgesehen Nebulisierung. Der Grund dafür ist, dass zu dem damaligen Zeitpunkt ein entsprechender Aufbau bzw. Möglichkeit zur Nebulisierung noch nicht vorhanden war und die Fragmentierung mit Ultraschall bei früheren Projekten erfolgreich durchgeführt wurde [124,125].

Die Größenverteilung der Fragmentierung der vier cDNAs nach einer SPRI-Aufreinigung wurde auf einem 7500 DNA-Chip gemessen, und wie dort ersichtlich ist, war die Fragmentierung nicht vollständig. Es wurden aber Fragmentierungsprodukte erzeugt, deren Menge auf dem Chip deutlich zu sehen ist (Messkurve zwischen 300 bp und 1500 bp). Aus früheren Versuchen war bekannt, dass man mit diesen Mengen am Ende der Library-Produktion deutlich messbare und für folgende emPCRs ausreichend konzentrierte Librarys erhalten konnte, daher wurde auf einen erneuten Fragmentierungsschritt über Ultraschall verzichtet. Es ist auch fraglich, ob eine deutliche Verlängerung der Zeit im Ultraschall zu einer letztendlich höheren Ausbeute an verwendbaren Librarys führen würde; Library-Produktionen bei anderen Projekten haben gezeigt, dass dies nicht der Fall ist, was vermutlich an einer Häufung von Einzelstrangbrüchen bei kleinen Fragmenten liegt, die während der Library-Produktion (Nick-Reparatur) nicht vollständig abgefangen werden können und zu nicht amplifizierbaren „Pseudo“-Library-DNAs führen (Insert DNAs mit nur Adapter B-Sequenz), was sich in stark erhöhten cpb-Zahlen ausdrückt (ist ein Teil der gemessenen Library nicht amplifizierbar, so muss eine deutlich größere Menge an Library eingesetzt werden, die cpb-Zahl steigt), weshalb keine längere Fragmentierungszeit gewählt wurde. Bei der Berechnung von Library-Konzentration und der emPCR-Ansätze ist später darauf zu achten, eine entsprechende durchschnittliche Länge bei den vier cDNAs zu verwenden. Betrachtet man den Messkurvenverlauf der fragmentierten cDNAs auf dem 7500 DNA-Chip, wird auch deutlich, dass die SPRI-Aufreinigung erfolgreich war, da es keine Fragmente kleiner als 300 bp gibt und die Fragmente, die kleiner als 500 bp sind, deutlich verringert wurden. Da die cDNAs nach Fragmentierung und SPRI-Aufreinigung also in ausreichender Menge vorlagen und von kleinen Fragmenten befreit waren, wurde sie für die Library-Herstellung weiter verwendet.

Es folgten dem Protokoll entsprechend Endrepair, Ligation, eine weitere SPRI-Aufreinigung, Nick-Reparatur und zuletzt die Elution der einzelsträngigen 454-cDNA-Librarys. Die Messung der Größenverteilung der so hergestellten Librarys erfolgte auf einem RNA Pico 6000 Chip. Auf diesem ist deutlich zu sehen, dass die vier aqf-cDNA-Librarys in einem entsprechenden Größenbereich vorhanden sind. Das wiederum bestätigt auch die aus früheren Library-Präparationen gemachten Erfahrungen und die Annahme, dass die auf dem 7500 DNA-Chip gemessene Menge an fragmentierter DNA für eine weitere Library-Herstellung ausreicht. Durch die weitere SPRI-Aufreinigung konnten die Fragmentierungsprodukte, die kleiner als 500 bp sind, im Vergleich zum 7500 DNA Chip weiter reduziert werden, sie sind kaum vorhanden. Bei den beiden cDNA-Librarys aqf3b und aqf4b sind zwei Peaks von kleinen Fragmenten sichtbar, die auf dem 7500 DNA-Chip nicht auftraten. Handelt es sich dabei um echte DNA-Fragmente und nicht um Messfehler, so müssen sie bei einem auf Fragmentierung und erste SPRI-Aufreinigung folgenden Library-Produktions-Schritt

eingeführt worden sein. Sind das echte Kontaminationen, ohne Adapter-A- und –B-Sequenzen, so würden sie bei dem weiteren Versuchsverlauf (emPCR, Sequencing) weitgehend folgenlos sein und lediglich die Konzentrationsberechnung der Library etwas verzerren. Auf einen Sequencingrun hätten sie keine Auswirkung, da bei Fehlen von Adapter-A- und –B-Sequenzen eine Amplifikation in der emPCR und ein Sequenzieren nicht möglich ist. Als weitere Möglichkeit könnte es sich bei den kleinen Fragmenten um einzelne unverbrauchte Adapter oder Adapterdimere handeln, die die zweite SPRI-Aufreinigung überstanden haben. Sind es einzelne Adapter oder die Adapterdimere B-B oder A-A, so sollten sie ebenfalls keine Auswirkungen auf den weiteren Verlauf in emPCR und Sequencing haben, abgesehen von einer leichten Verzerrung bei der Konzentrationsberechnung –da die Librarys vor einem großen Sequencing-Run-Start titriert werden, um die beste cpb-Zahl herauszufinden, würde das so abgefangen werden. Möglicherweise problematisch könnten die Adapterdimere mit der Sequenz A-B (oder B-A) werden oder eine wenige Basen lange Kontamination (1 bis ca.10 bp) mit Adapter-A- und –B-Sequenzen, da diese sich in einer emPCR erfolgreich amplifizieren lassen würden und dementsprechend ins Sequencing gehen (und auch dort aufgrund der A-Sequenz sequenzierbar sind). Da einerseits nicht feststand, ob die kleinen Fragmente bei aq3b und aqf4b tatsächlich A-B-Konstrukte sind und andererseits die Konzentrationsverhältnisse zwischen kleinen Peaks und cDNA deutlich dafür sprechen, dass, selbst wenn es sich bei den kleinen Peaks um sequenzierbare Produkte handelt, diese keine starken Auswirkungen auf emPCR und Sequencing haben, wurden alle vier aqf-Librarys für eine weitere Verwendung freigegeben.

Die durchschnittliche Länge wurde bei allen vier Librarys anhand der DNA- und RNA-Chips auf 700 bp geschätzt. Ob dies tatsächlich so ist, lässt sich anhand der Chips nicht herausfinden. Um das bestimmen zu wollen, müsste man die Größenverteilung aller 454-Sequencing-„aktiven“ DNA-Stränge kennen, was nicht möglich ist. 454-Sequencing-„aktive“-DNA-Stränge sind die Teile der Librarys, die die Adapter-A- und –B-Sequenzen tragen, was nicht notwendigerweise der Fall sein muss, da auch bei einer theoretisch vollständig erfolgreich aufgereinigten Library aufgrund des General Library-Protokolls immernoch DNA-Stränge in der Library vorhanden sein können, die eine oder zwei Adapter-B-Sequenzen tragen. Diese werden auf dem RNA-Chip ebenfalls mitgemessen. Dazu kommt noch, dass sich nicht notwendigerweise alle DNA-Stränge in einer emPCR gleichmäßig amplifizieren lassen; neben der schlechteren Amplifikation längerer Stränge können auch Amplifikationsunterschiede, die sich aufgrund der Sequenzen selber ergeben, nicht ausgeschlossen werden (schwierige Sequenzbereiche, Ausbildung von Sekundärstrukturen, Loops, etc.). Aufgrund dieser Überlegungen ist eine Schätzung zwar wichtig, da sie für die emPCR nötig ist, aber kann nicht exakt sein. Die genaue bzw. ideale Einstellung der Librarys für die emPCR wird praktisch über eine Titration ermittelt; dafür sollte die Schätzung in dem richtigen Bereich liegen, Unterschiede von beispielsweise wenigen 100 bp können und werden über die Einstellung des cpb-Wertes abgefangen.

4.2.1.2 Titration der General Librarys

Mit den vier General Librarys wurden Titrationen mit cpb-Zahlen von 0,5, 1, 2 und 4 durchgeführt. Es wurden diese cpb-Werte verwendet, weil sie von Roche empfohlen [85] und damit bei vorangegangenen Projekten in der Mehrzahl gute Erfahrungen bei der Titration von General Librarys gemacht wurden.

Zur Titration wird dabei von Roche das GS FLX Titanium SV emPCR Kit (Lib-L) verwendet. Dieses kam bei den durchgeführten Titrationsen nicht zur Anwendung. Stattdessen wurden die benötigten Mengen und Volumina an Enzymen und Reagenzien auf ein kleineres Titrationsvolumen heruntergerechnet und das GS FLX Titanium LV emPCR Kit (Lib-L) benutzt. Das geschah aus Kostengründen, da beide Kits den gleichen Preis hatten, aber das LV-Kit die doppelte Reaktionsmenge besitzt [85]: ein SV-Kit besitzt 16 Tubes für einzelne emPCR-Ansätze, also 16 Reaktionen insgesamt, wobei für eine Titration mit vier verschiedenen cpb-Werten je ein Tube, also insgesamt vier verwendet werden. Das LV-Kit besitzt eine Menge von 32 Reaktionen. Da sich auch kleinere Mengen als die von einer Reaktion gut im Laboralltag bearbeiten lassen, wurden die Mengen des LV-Kits nicht auf 1 Reaktion, sondern auf 0,5 Reaktionen für einen cpb-Wert heruntergerechnet (weshalb 1/64 aller Mengen und Volumina verwendet wurden), um zusätzlich Kosten zu sparen. Der einzige Unterschied bei der Handhabung betrifft dabei das Emulsions-Öl; es muss in kleinere Tubes aliquotiert werden, beim SV-Kit liegt es schon in kleineren Tubes vor. Alle anderen Reagenzien müssen bei beiden Kits in entsprechende Mengen aliquotiert werden (Capture-Beads, PCR-Mix, usw.), so dass es keine Unterschiede zwischen einem SV-Kit und einem heruntergerechneten LV-Kit beim Pipettieren und damit verbundenen Fehlern gibt (z.B. bei der Capture-Bead-Menge, die wichtig für eine Auswertung der Titration ist).

Die Ergebnisse der Titration lassen nur bei einer Library, aqf5b, die Bestimmung eines geeigneten cpb-Wertes alleine aus den Messwerten zu; ein cpb-Wert von 0,5 führte in der Titration zu einer Enrichment-rate von rund 8%. Das ist groß genug für eine ausreichende Anzahl an Beads und liegt deutlich unter der von Roche vorgegebenen Obergrenze von 15-20%, so dass in einer emPCR von aqf5b mit 0,5 cpb der überwiegende Teil von DNA-Beads monoklonal sein sollte und gute Ergebnisse im Sequencing liefern müsste.

Die bestimmten Enrichment-Rates bei den drei anderen Librarys lagen bei allen vier cpb-Werten innerhalb oder über dem von Roche vorgegebenen oberen Grenzwert von 15-20%. Die Enrichment-Rates von aqf3b und aqf4b waren bei 0,5 cpb mit 17,4% bzw. 19,9% innerhalb des Grenzbereiches. Um eine möglichst große Anzahl monoklonaler Beads zu erhalten, sollten etwas kleinere cpb-Werte zur Anwendung kommen, weshalb für aqf3b 0,4 cpb und für aqf4b 0,35 cpb gewählt wurden. Mit zunehmender DNA-Menge, die zu einer gleichen Anzahl Capture-Beads gegeben wird, sollte auch die Anzahl an Capture-Beads steigen, die DNA bindet und an denen DNA in der emPCR amplifiziert wird. Also sollte die Enrichment-Rate prinzipiell steigen, je größer die zugegebene DNA-Menge ist -wobei es bei großen DNA-Mengen zu einer Sättigung kommen muss, da auf der einen Seite nur eine begrenzte Anzahl an Capture-Beads vorhanden ist und auf der anderen Seite nicht nur ein DNA-Strang binden kann, sondern auch mehrere an einen Capture-Bead. Dieser Verlauf lässt sich anhand der Titrationsergebnisse erahnen; aqf6b scheint sich bei zwei und vier cpb schon nahe dem Sättigungsbereich (von 100%) zu befinden, die drei anderen Librarys sind davon noch entfernt, da die Enrichment-Rates auch bei 4 cpb unter 50% bleiben. Würde man keine DNA zu den Capture-Beads geben, so dürften auch keine im Beaderecovery aufgereinigt werden, die Enrichment-Rate muss Null sein. Daher müssen alle Titrationskurven bei Null cpb eine Enrichment-Rate gleich Null haben. Anhand dieser Überlegung wurde versucht, für aqf6b einen geeigneten cpb-Wert für eine gewünschte

Enrichment-Rate von 8-10% zu finden. Es wurde ein linearer Verlauf zwischen dem Nullpunkt und dem Messwert bei 0,5 cpb für aqf6b angenommen und so ein cpb-Wert von 0,05 geschätzt. Die Titration sollte nur für aqf6b alleine nicht wiederholt werden; in einer emPCR für den folgenden Sequencing-Run würde man aqf6b in geringerem Maß zugeben als die anderen drei Librarys, um einen möglichen größeren Fehler bei der cpb-Wahl abzufangen.

Vergleicht man abschließend bei aqf3b, aqf4b und aqf5b die erhaltenen cpb-Werte für Enrichment-Rates von 8-10%, so liegen diese nicht sehr weit auseinander (0,35 bis 0,5). Sollte es sich bei den auf dem RNA-Chip sichtbaren Peaks bei aqf3b und aqf4b im Bereich unterhalb 25s tatsächlich um DNA-Fragmente und nicht um Messfehler handeln, so hatten diese bei der Titration keinen oder kaum Einfluss auf die Amplifikation. Amplifizierbare Fragmente von einer Größe zwischen 60 und 100 bp hätten zu deutlich höheren Enrichment-Rates bei vergleichbaren cpb-Werten führen müssen, da sie aufgrund ihrer kurzen Sequenz und den auf dem RNA-Chip sichtbaren Peaks in einer hohen molaren Konzentration vorlägen. Bei aqf6b hingegen lässt sich das nach den Ergebnissen der Titration nicht ausschließen –die Messung auf dem RNA-Chip bietet keine Aussage über mögliche kurze Fragmente und die hohen Enrichment-Rates könnten entweder kleinere Fragmente vermuten lassen, oder die bei aqf6b gemessene Konzentration war zu niedrig und ist in Wirklichkeit höher.

4.2.1.3 Beadrecovery der General Librarys

Die vier aqf-General-Librarys wurden für eine große emPCR auf zwei Cups verteilt, wobei aqf3b und aqf4b in einem Cup und aqf5b und aqf6b in einem anderen Cup pipettiert wurden. D.h. zwei Librarys teilen sich ein Aliquot Capture-Beads. Nach durchgeführter emPCR wurde diese aufgereinigt. Mit Beadrecovery-Rates beider Cups von 83,9 bzw. 81,4% liegt man kurz unterhalb der von Roche als „typisch“ bezeichneten 85%. Bei der Durchführung des Recovery können Beads vorallendingen bei der Emulsions-Kollektion durch eine zu stark eingestellte Vakuumpumpe und beim Abgießen der Überstände nach den ersten Zentrifugationsschritten verloren gehen; außerdem verbleiben immer ein paar Rückstände in den Schläuchen der zu Emulsionskollektion verwendeten Saugvorrichtung. Daher sind die erreichten Ausbeuten an Beads gut, da sie nur knapp unterhalb der beschriebenen 85% liegen; außerdem hat es sich im Laboralltag bei Beadrecoverys anderer Projekte gezeigt, dass es generell schwer ist, Ausbeuten von mehr als 85% zu erreichen.

Die Enrichment-Rates beider Cups liegen unterhalb des Grenzbereiches von 15-20%; allerdings weichen sie von den aus der Titration erwarteten Raten deutlich ab. In beiden Cups wurden jeweils 50% der Beads auf eine Library berechnet, d.h. die Librarys sollten in gleichen Anteilen vorliegen. Laut Titration sollte aqf3b bei dem verwendeten cpb-Wert von 0,4 und aqf4b bei einem cpb-Wert von 0,35 zu Enrichment-Rates jeweils zwischen 8 bis 10% führen. Beide Librarys wurden auf jeweils die Hälfte der in dem Cup A verwendeten Capture-Beads berechnet, so dass es theoretisch bei Cup A zu einer Enrichment-Rate von ebenfalls 8 bis 10% anstelle der gemessenen 13,9% hätte kommen sollen. Für Cup B gilt das gleiche, nur dass hier die Abweichung zwischen dem erwarteten Ergebnis von 8 bis 10% und der Messung von 2,1% größer ist. Die bei beiden Cups auftretenden Abweichungen können mehrere Ursachen haben: bis auf aqf5b mussten alle anderen cpb-Werte bei der Titration geschätzt werden, vielleicht lag die Schätzung zu weit von der Realität entfernt; außerdem könnte die Titration selber innerhalb der gemessenen Werte fehlerbehaftet oder für die größeren Volumina bei dem

Ansatz der großen emPCR nicht aussagekräftig sein. Geht man davon aus, dass die Titration selber aussagekräftig ist innerhalb der gemessenen Werte und die Schätzung fehlerbehaftet, dann ist das Ergebnis von Cup B mit einer Enrichment-Rate von 2,1% zu niedrig; selbst wenn man annimmt, dass aqf6b nicht amplifiziert wurde, weil man mit der Schätzung eines cpb-Wertes für Enrichment-Rates zwischen 8 bis 10% bei der Titration vollkommen daneben lag, dann müsste die Enrichment-Rate von Cup B trotzdem bei rund 4% liegen, da die Hälfte der verwendeten DNA von aqf5b stammt, der in der Titration bei 0,5 cpb Enrichment-Rates von 7,9% aufwies.

Die Enrichment-Rates lassen auch nur Rückschlüsse auf die gesamte Beadmenge zu, wie die einzelnen Librarys auf die Beads verteilt sind, kann so nicht bestimmt werden. Da die Enrichment-Rates beider Cups trotz der teils starken Abweichungen der Messwerte zu den Erwartungen in einem Bereich liegen, der gut sequenzierbare Beads anzeigt, wurden diese Beads für einen Laufstart weiterverwendet.

4.2.1.4 Sequencing Run der General Librarys.

Mit den aus dem Beadrecovery erhaltenen DNA-Beads wurde ein 454-Sequencing-Lauf gestartet. Mit einer Gesamtmenge von 478867 Reads und einer durchschnittlichen Länge von 280 bp der Reads lieferte der Sequencing Run ein schlechtes Ergebnis; bei einem guten Run erhält man mindestens um die 800000 Reads und durchschnittliche Readlängen um die 350 bis 400 bp. Zu schlechten Sequencing-Run-Ergebnissen kann eine Vielzahl von Ursachen führen: die Librarys können problematisch sein, in emPCR und Beadrecovery können Fehler auftreten, ebenso beim Sequencing selber; ferner kann es Probleme mit dem Sequenzier-Gerät geben, dieses kann fehlerhaft arbeiten. Außerdem können die verwendeten Reagenzien von schlechterer Qualität sein.

Ein Gerätefehler ist sehr unwahrscheinlich, da der Sequenzer zu dem Zeitpunkt bei vorangegangenen Runs und später erfolgten Runs gute Ergebnisse lieferte; das Sequencing-Tray, also der Behälter, der die Sequenzier-Reagenzien enthält, sah nach dem Run unauffällig aus. Die Reagenzien waren nach dem Run in üblichen Restmengen vorhanden, womit auch Verstopfungen der Filtereinheiten auf den Sipper-Tubes, die bei jedem neuen Run gewechselt werden, ausgeschlossen werden können. Kann man Fehler am Gerät ausschließen, was hier der Fall ist, so müssen die Probleme an anderer Stelle aufgetreten sein. Auch wenn die Librarys von der Titration abweichende Ergebnisse bei den Enrichment-Rates in der großen emPCR-lieferten, so waren diese „unauffällig“ und innerhalb der Vorgaben; lediglich das Verhältnis der einzelnen Librarys untereinander hätte variieren können. Die emPCR und das darauf erfolgende Beadrecovery sollten also nicht Ursache des schlechten Sequencing-Runs sein.

Die geringe Ausbeute an Reads kann auch nicht mit einer fehlerhaften Beladung der PTP oder einem Versagen des vorangegangenen Beadrecoverys erklärt werden: es wurden auf beiden Regionen Raw-Wells in einer Anzahl von knapp unterhalb 1 Mio. gefunden, was sehr nah an die Vorgaben von Roche für einen guten Run kommt (1-1,2 Mio.). Es wurden also weder beim Beladen der PTP DNA-tragende Beads herausgespült beim Abziehen der Layer, noch war das Beadrecovery fehlerbehaftet und es wurde eine große Fraktion von Beads mitgeschleppt, die keine Amplifikationsprodukte tragen (um als Rawwell erkannt zu werden, muss eine Signalgeneration erfolgen).

Um das Sequenzieren alleine bewerten zu können, muss man sich die Ergebnisse der Kontroll-DNA betrachten. Da die Kontroll-Beads erst bei der Beladung der PTP dazugegeben werden, durchlaufen sie die vorangegangene emPCR und das Beadrecovery nicht. Über die Kontroll-DNA lassen sich also Rückschlüsse auf die Qualität des Sequencings (Gerät) und der Sequencing-Reagents ziehen. Mit einem Passed Filter Well-Wert von 82% liegt die Kontroll-DNA in einem Bereich, der noch als gut bewertet werden kann. Bei Problemen mit den Sequenzier-Reagenzien (oder Geräte-Fehlern wie Pumpenproblemen) bricht der Anteil an Passed Filter Wells ein, liegt weit unterhalb von 50%. Schlechter als bei einem guten Run dagegen ist die durchschnittliche Länge der Kontroll-DNA, sie ist um rund 70 bp kürzer. Auffällig ist auch, dass die Qualität der erzeugten Kontroll-DNA-Reads schlechter ist -der Anteil an Kontroll-DNA-Reads mit Übereinstimmungen von 95%, 98% und 100% bei allen Längenbereichen ist geringer, fällt aber auch nicht so schlecht aus wie bei Runs, bei denen ein Geräte-Fehler vorlag; dort sind die Anteile selbst bei 95% Übereinstimmung viel geringer (sie können unter 20% fallen). Zusammenfassend kann man sagen, dass die Kontroll-DNA-Reads zwar in ausreichendem Anteil (Passed Filter Wells) vorhanden sind, aber ihre Qualität (Länge und Anteil bei den Übereinstimmungen) ist schlecht. Ein Problem bei den Sequencing-Reagenzien scheidet als generelle Ursache für den schlechten Lauf aus.

Die kürzeren durchschnittlichen Leseweiten der Kontroll-DNA und besonders auch die schlechteren Anteile bei den Übereinstimmungen von 95%, 98% und 100% können durch Überstrahlung aus Nachbar-Wells zustande kommen: sind dort DNA-Stränge an den Beads, die besonders viele homopolymere Bereiche tragen, so führt das zu starken Signalen, die die Signale anderer benachbarter Wells überstrahlen bzw. falschpositive Signale erzeugen können. Dadurch wäre eine Verschlechterung der Leseweiten der Kontroll-DNA und die schlechteren Werte bei der Übereinstimmung zu erklären. Und genau für derartige möglichen Überstrahlungen gibt es bei diesem Run Hinweise, wodurch das schlechte Run-Ergebnis zu erklären ist: der Anteil an Reads, die durch den Mixed Filter entfernt wurden, ist besonders hoch, die absoluten Zahlen ebenfalls, insgesamt 727896 Reads wurden so entfernt. Der Mixed Filter entfernt Reads aus Wells, bei denen keine eindeutige Signal-Zuordnung möglich ist, weil sie nicht monoklonal sind aufgrund von mehr als einem Amplifikationsprodukt pro Well -oder aber aufgrund von starken Überstrahlungen. Wegen der guten und niedrigen Enrichment-Rates des Beadrecovery kann ersteres ausgeschlossen werden. Mit den Überstrahlungen können die schlechteren Ergebnisse der Kontroll-DNA erklärt werden, und der schlechte Run insgesamt. Bei der Durchsicht der Fasta-Files durch Sven Klages fielen viele kurze Sequenzen mit homopolymeren Bereichen auf, besonders Poly-T, nach denen die Reads abbrachen. Diese homopolymeren Bereiche erzeugen hohe Signalintensitäten und werden die Ursache für mögliche Überstrahlungen und damit für den schlechten Run sein.

Das Auftreten besonders der Poly-T Regionen am Anfang der Sequenzen lässt sich anhand der Library-Herstellung erklären: die vier aqf-cDNAs tragen alle einen Poly-A-Schwanz am 3'-Ende, bzw. damit ein Poly-T am 5'-Ende des Gegenstranges und wurden fragmentiert, weil sie zu lang sind. Verbunden mit der bei der General-Library-Herstellung eingesetzten Random-Ligation kommt es zu Library-Konstrukten, die nach der Adapter-A-Sequenz mit dem Poly-T-Bereich beginnen. Und die Beads dieser Konstrukte sind es, die ein starkes Signal gleich zu Anfang des Sequencing geben, was

zu Überstrahlungen und damit verbundenen hohen Mixed-Filter-Ausschlüssen führt und zu der Verschlechterung der Kontroll-DNA-Sequenzen.

Es stellt sich demnach die Frage, ob diese Art der 454-Library-Herstellung überhaupt geeignet für die Sequenzierung der aqf-cDNAs ist bzw. von cDNA-Konstrukten, die ein Poly-A am 3'-Ende besitzen. Auch wenn man versuchen würde, niedrigere Enrichment-Rates anzustreben durch einen etwas geringeren cpb-Wert bei Ansatz der emPCR –Region 1 ist etwas besser gewesen als Region 2 im Sequencing, da erstere aus rund 27% Beads von Cup B besteht, der mit 2,1% eine deutlich geringere Enrichment-Rate hatte als Cup A mit rund 14%, so würden die Ergebnisse im Sequencing sich etwas verbessern lassen, analog zum Verlauf von Region 1, aber sie blieben damit insgesamt immernoch schlecht; ferner ändert das nichts an der Problematik, dass aufgrund der Random-Ligation eine Vielzahl von Librarys mit Poly-T starten.

Mit den vorhandenen General-aqf-Librarys können also keine weiteren Sequenzierungen durchgeführt werden, weil die Ausbeuten zu gering und damit die Kosten zu hoch sind. Auch Änderungen des General-Library-Protokolls erscheinen fragwürdig; man könnte versuchen, die Fragmentierungszeit zu verlängern, umso mehr Strangbrüche der cDNAs zu erreichen. Allerdings wären die Aussichten auf Erfolg gering, da die cDNAs in einem Größenbereich von 500 bp bis 2000 bp liegen; die meisten Stränge, die mehr als einmal brechen, würden aufgrund ihrer geringen Größe bei der notwendigen SPRI-Aufreinigung entfernt werden.

4.2.2 Sequenzierung von aqf-cDNAs über die Erstellung von Amplicon Librarys

Wie aus den vorhergegangenen Versuchs- und Runergebnissen deutlich wurde, stellt der Poly-A-Schwanz der cDNAs ein großes Problem beim Sequenzieren dar. In einem zweiten Anlauf wurde versucht, aus den cDNAs für die 454-Sequenzierung Amplicon Librarys zu erstellen.

Der Nachteil liegt dabei darin, dass Amplicon Librarys auf dem 454 etwas schlechtere Run-Ergebnisse liefern als General-Librarys (Shotgun-Librarys); bei guten Amplicon-Läufen kann man insgesamt ca. 750000 Reads erwarten [94], was unterhalb der 800000-1,2 Mio. von Shotgun-Librarys liegt. Der Vorteil bei der Erstellung einer Amplicon Library gegenüber einer Shotgun-Library ist die Festlegung der Sequenzrichtung; die 454-Adapter-Sequenzen werden nicht über Random-Ligation an die Insert-DNA gebracht, sondern über eine PCR. Damit kann festgelegt werden, aus welcher Richtung man das Template sequenziert. Und da der Poly-A-Schwanz bei den aqf-cDNA ein Problem ist, können bei der Erstellung einer Amplicon Library die Primer so gewählt werden, dass sich der Poly-Schwanz immer am 3'-Ende der Sequenzierprodukte befindet.

4.2.2.1 Erstellung der Amplicon Librarys

Aus den sechs cDNAs aqf3b, aqf4b, aqf5b, aqf6b, aqf9b und aqf10b wurden in einer Gradienten-PCR Amplicon Librarys erstellt. Jeder Library wurde ein anderes MID über den Forwardprimer gegeben. Als Revers-Primer wurden zwei verschiedene Varianten ausprobiert: der ADB-CDNA-Revers-Primer, der sowohl mit zehn Basen in den Poly-A-Bereich hereingeht als auch noch an neun Basen der MINT-Sequenzen bindet; und der ADB-TANKER-Revers-Primer, der ausschließlich den Poly-A-Bereich bindet. Damit ist die Sequenzierrichtung vorgegeben, in den Poly-A-Bereich der cDNAs kann somit nur hineingelesen werden, wenn das Sequenzier-Enzym überhaupt bis zum Ende der Library durchliest, eine Sequenzierung des revers komplementären Poly-T gleich zu Beginn ist ausgeschlossen. Anhand der Ergebnisse der Gradienten-PCR sollte entschieden werden, welches Amplifikationsprodukt für die Sequenzierung eingesetzt wird. Wie die Ergebnisse der Gelbilder zeigen, verlief die PCR erfolgreich, es wurde Produkt in den erwarteten Größen der ursprünglichen cDNAs hergestellt und die Negativkontrollen waren frei von Amplifikationsprodukten. Es mussten zwei Dinge entschieden werden: a) mit welchem Revers-Primer wurde die „bessere“ Library erzeugt und b) bei welcher Annealing-Temperatur wird das entsprechende Produkt weiterverwendet zum Sequenzieren.

Beide Revers-Primer liefern bei gleicher Annealing-Temperatur Produkte mit gleicher Länge, allerdings scheinen die ADB-TANKER-Produkte etwas schwächer bei allen Librarys zu sein. Aus diesem Grund wurde sich gegen eine Weiterverarbeitung der TANKER-Produkte entschieden, da bei dem anderen Revers-Primer die Ausbeuten etwas größer sind. Außerdem kann der TANKER-Revers-Primer aufgrund seiner Sequenz etwas anfälliger für Annealing-Fehler sein, da am 3'-Ende mehr Poly-T vorhanden sind als bei dem ADB-CDNA-Primer; dieser bindet noch mit einem GC-reichen Bereich an die cDNA. Bei höheren Annealing-Temperaturen sollte die Spezifität der Basenpaarung bei diesem besser sein als beim TANKER-Revers-Primer.

Für die Entscheidung, welche Annealing-Temperatur am geeignetsten ist und demnach welcher Library-Ansatz weiterverwendet werden soll, muss man den Verlauf der Größenverteilung der Amplifikations-Produkte betrachten. Bei zu niedrigen Annealing-Temperaturen kann es zu unspezifischen Amplifikations-Produkten kommen, da die Primer Fehlpaarungen eingehen können. Bei zu hohen Temperaturen findet kein Annealing mehr und dementsprechend keine Produkt-Amplifikation mehr statt. Letzteres tritt innerhalb des verwendeten Temperaturbereiches von 50 bis 70°C nicht auf, wenn auch die Ausbeuten zu Position 12 hin (70°C) deutlich kleiner werden. Bei allen Temperaturen sind die Bereiche zwischen 0,5 und 2 kb vertreten, allerdings scheint sich der Schwerpunkt des DNA-Schmiers mit steigender Annealing-Temperatur nach unten zu kleineren Amplifikaten zu verlagern. Bei allen Librarys kann man ab Position 6 keine Veränderung zur Position 12 hin mehr feststellen was die Größenverteilung angeht, lediglich die Ausbeuten sinken etwas. Davon ausgehend, dass eine höhere Annealing-Temperatur eine höhere Spezifität bedeutet, wurden deshalb die Amplifikations-Produkte auf Position 6 mit dem ADB-CDNA-Revers-Primer von allen Librarys ausgewählt und weiterverarbeitet.

Wie anhand der Gelbilder deutlich wird, so treten bei allen Temperaturen Amplifikations-Produkte auf, die kleiner als 500 bp sind, dementsprechend auch in den Ansätzen der Position 6. Daher erfolgte eine zweimalige Ampure-Bead-Aufreinigung mit Größen-Cutoff bei 500 bp, um diese loszuwerden. Die

Ampure-Bead-Aufreinigung wurde zweimal ausgeführt, da genug Material für die Aufreinigung vorhanden war und sicher gestellt werden sollte, dass die Amplicon Librarys frei von kleiner Library-DNA sind und um die Amplicon Librarys so gut es geht mit den General-Librarys vergleichen zu können; bei diesen wird auch insgesamt zweimal mit SPRI aufgereinigt. Für die weitere Verwendung musste noch eine durchschnittliche Größe der Amplicon Librarys festgesetzt werden (zur Berechnung der emPCR). Da die Amplifikations-Produkte bei Position 6 zwischen 500 bp und 2 kb liegen (nach den SPRI-Aufreinigungen sind Längen unter 500 bp größtenteils weggereinigt), wurde eine durchschnittliche Länge von 1 kb für alle Librarys festgelegt. Es ist unerheblich, ob das tatsächlich genau dem Maximum der Längenverteilung entspricht oder nicht, wichtig ist nur, dass die geschätzten 1 kb durchschnittliche Länge dem richtigen Wert nahe kommt. Wie auch bei den General-Librarys kann nicht festgestellt werden, wie die Konzentration und die durchschnittliche Länge der 454-„aktiven“-Librarymoleküle ist. Die genaue Einstellung der Library muss über eine Titration erfolgen, bei der die cpb-Zahl als „Korrekturfaktor“ bestimmt wird. Für eine erfolgreiche Titration bzw. eine leichtere Interpretation der Titrationsergebnisse sollten die DNA-Konzentrationen und geschätzten durchschnittlichen Längen zumindest nicht in Zehnerpotenzen von den tatsächlichen Werten abweichen; Fehler bis beispielsweise 40-50% können von der Titration abgefangen werden.

4.2.2.2 Titration der Amplicon Librarys

Es wurde eine Titration mit cpb-Werten von 0,5, 1, 2 und 4 mit den sechs Amplicon Librarys durchgeführt. Die Ergebnisse der Titration lassen bei fast allen Librarys eine Bestimmung eines geeigneten cpb-Wertes anhand der Messwerte zu. Es wurden bei aqf3b, aqf4b, aqf6b, aqf9b und aqf10b cpb-Werte ausgesucht, die zu einer Enrichment-Rate von rund 10% führten. Dieser Wert weist auf gute Beads für das Sequencing hin, und bietet Spielräume nach oben und unten, sollte es bei den großen emPCRs zu Abweichungen der Enrichment-Rates kommen, weil sich entweder dort kleine Fehler einschleichen oder die durchgeführten Titrations fehlerhaft waren und die eigentlichen Enrichment-Rates etwas davon abweichen. So sind Fehler von bis zu 5% Enrichment-Rate tolerierbar, da bei 5% zwar etwas zu wenig aber immer noch hinreichend viele Beads für einen Runstart vorhanden sein müssten, und eine Enrichment-Rate von 15% gerade noch vor dem von Roche angegebenen oberen Grenzbereich von 15-20% liegt.

Nur bei aqf5b musste ein passender cpb-Wert geschätzt werden, da die Enrichment-Rates innerhalb der verwendeten cpb-Zahlen etwas zu niedrig waren. Dabei wurde angenommen, dass die Beadrecovery-Rate von 4 cpb zu hoch und die sich daraus ergebende Enrichment-Rate zu niedrig ist. Da nicht entschieden werden konnte, ob die Enrichment-Rate bei 0,5 cpb zu hoch oder die bei 1 cpb zu niedrig ist, um einen Anstieg zu 2 und dann 4 cpb zu erhalten, wurde für aqf5b ein cpb-Wert von 6 geschätzt, da hier in beiden Fällen vertretbare Enrichment-Rates zu erwarten sind (zwischen ca. 10 und 15%).

Da die Ergebnisse der Titration bei allen Librarys im Großen und Ganzen innerhalb des erwarteten Bereichs liegen –bei keiner der Librarys liegen sie für alle verwendeten cpb-Werte in einem Bereich von über 20% oder unter 3%, sollten die geschätzten durchschnittlichen Längen der Realität nahe kommen.

4.2.2.3 Beadrecovery der Amplicon Librarys

Die sechs Amplicon Librarys wurden in einer großen emPCR amplifiziert, wobei jede Library separat angesetzt wurde. Dafür wurden die entsprechenden Mengen und Volumina der in der emPCR verwendeten Enzyme, Reagenzien und Puffer auf ein Sechstel heruntergerechnet. Damit sollte sichergestellt werden, dass es zu keinen Problemen bei der Amplifikation kommt durch Wechselwirkung mit anderer Library-DNA (ungleichmäßige Verteilung auf die Beads, eine Library wird verdrängt, usw.), es sollten ideale Bedingungen herrschen.

Die Beadrecovery-Rates liegen dicht an der von Roche als „typisch“ bezeichneten Rate von 85%, d.h. es ging nur ein geringer Teil an Beads während der Rückgewinnung bei der Durchführung verloren (Abgießen der Überstände, Verbleiben in Schläuchen). Auch bei diesem Recovery hat es sich gezeigt, dass es generell schwer ist, Raten über 85% zu erreichen. Die Enrichment-Rates aller Librarys mit Ausnahme von aqf9b liegen unterhalb des Grenzbereiches von 15-20%; die angereicherten Beads sollten demnach also gut sequenzierbar sein. Auch das Beladen eines Sequencing-Runs mit Beads von aqf9b ist vertretbar, da dessen Enrichment-Rate immer noch im Grenzbereich liegt und mit 16,9% nicht an der Obergrenze ist, außerdem macht aqf9b nur ein Sechstel der gesamten Beadmenge aus bzw. ein Drittel auf einer Region, auf der diese Beads geladen werden könnten. Im Run wird sich zeigen, ob diese Beads überhaupt einen Einfluss auf die Ausbeuten der Region haben.

Bei drei Librarys lagen die Enrichment-Rates sehr nahe an Ergebnissen der Titration, aqf5b, aqf6b und aqf10b lieferten Raten zwischen 9 und 10%. Die drei anderen Librarys waren mit Enrichment-Rates von 5,7% (aqf3b), 14,5% (aqf4b) und 16,8% (aqf9b) rund 5% von den Erwartungen anhand der Titrations-Ergebnissen entfernt (10%). Das zeigt, dass anhand der Titration eine Einschätzung möglich ist, diese aber doch Abweichungen bei den großen emPCRs unterworfen zu sein scheint.

4.2.2.4 Sequencing Run der Amplicon Librarys

Mit den aqf-Amplicon Library tragenden Beads wurde ein Sequencing-Run gestartet. Mit einer Ausbeute von insgesamt nur 163305 Reads bei einer durchschnittlichen Leselänge von 150 bp ist dieser Sequencing-Run als sehr schlecht zu werten. Ein guter Amplicon-Lauf sollte um die 750000 Reads bringen und höhere, der Library-Länge entsprechende Leseweiten. Ein Gerätefehler scheidet auch wie bei dem Run der General-Librarys aus, da Runs mit anderen Projekten vor und nach diesem Amplicon-Lauf fehlerfrei und gut verliefen, es gab nach dem Run selber auch keinerlei Auffälligkeiten am Gerät (ausgelaufene Flüssigkeiten, zu hohe Restvolumina der Sequenzier-Reagenzien).

Schaut man sich die Ergebnisse der emPCR bzw. des Beadrecoverys der aqf-Amplicon Librarys an, so waren die Enrichment-Rates in den vorgegebenen Bereichen und deuten eigentlich auf sequenzierbare Beads hin, die emPCR-Cups hatten Raten zwischen 5 und 10%, nur ein Cup (enthielt Beads der aqf9b-Library) lag mit 16,8% Enrichment-Rate etwas höher aber immer noch deutlich innerhalb des Grenzbereiches von 15-20%. Die Enrichment-Rates im Beadrecovery der großen emPCR lagen auch in den erwarteten Bereichen der Titrationsen und sind damit in zwei voneinander unabhängigen emPCRs erreicht worden, womit Fehlmessungen unwahrscheinlich sind. Die PTP wurde auch mit genügend sequenzierbaren Beads beladen, die Anzahl an erkannten Raw Wells ist mit 750000 (Region 1) bzw. 800000 (Region 2) ausreichend (wenn auch etwas unterhalb der

Vorgaben von 1 bis 1,2 Mio. pro Region –vielleicht gingen einige wenige Beads beim Abziehen der Layer bei der PTP-Vorbereitung verloren), d.h. die im Beadrecovery aufgereinigten Beads trugen Amplifikationsprodukte. Es gibt also keine Hinweise, dass emPCR und Beadrecovery fehlerbehaftet sind.

Anhand der Ergebnisse der Kontroll-DNA ist eine Bewertung des Sequencings unabhängig von emPCR, Beadrecovery und Library-DNA-Qualität möglich, da die Kontroll-DNA-Beads erst beim Sequencing dazugegeben werden. Bei diesem Run ist die Kontroll-DNA auf beiden Regionen unauffällig und weist auf ein fehlerfreies Sequencing hin: mit 82% Passed Filter Wells liegt die Kontroll-DNA in einem noch guten Bereich, die durchschnittliche Länge ist mit 400 bp (Region 1) bzw. 380 bp (Region 2) gut und folgt den Vorgaben für einen guten Run -letzteres war beim aqf-General-Library-Run anders, dort war die durchschnittliche Länge der Kontroll-DNA kleiner. Die Sequenziergenauigkeit der Kontroll-DNA-Reads liegt auch innerhalb der Vorgaben für einen erfolgreichen Run und ist ebenfalls höher als beim aqf-General-Library-Run –besonders über Längen von 200, 300 und 400 bp ist der Anteil an Kontroll-Reads, die zu 95 und 98% Übereinstimmung aufweisen, bei dem aqf-amplicon-Run um 10-20% höher als bei dem General-Library-Run. Es scheint demnach auch keine Probleme mit den Sequenzierreagenzien und dem Sequencing selber gegeben zu haben, sonst wären die Kontroll-DNA-Ergebnisse schlechter.

Schaut man sich die Anteile an Reads an, die herausgefiltert wurden, so fällt auf, dass mit 559324 Reads besonders viele als „Short“ und mit 429972 Reads als „Dots“ aussortiert wurden. Wirklich ungewöhnlich hoch ist auch die Zahl von 106634 Reads, die als „Failed Short Primer“ herausgefiltert wurden. Der Dot-Filter entfernt Reads mit zu vielen negativen Flows, also Flows, bei denen keine Base detektiert wurde. Das kann entweder an fehlendem (Read zu kurz) oder an einem zu schwachen Nukleotideinbau liegen, weil der nicht erkannt wird (das Signal selber ist zu schwach). Reads werden als „Failed Short“ durch den Signal Intensity-Filter und den TrimBack Valley-Filter heraussortiert. Der Signal Intensity-Filter entfernt Reads, deren Signale zwar erkannt aber als zu schwach gewertet werden; der TrimBack Valley-Filter entfernt Reads, die zu viele intermediäre Signale produzieren. Beide Filter entfernen also Reads, bei denen keine eindeutigen Signale zugeordnet werden können, z.B. aufgrund eines zu starken Hintergrundrauschens, weil die erzeugten Signale möglicherweise zu schwach sind. Die hohen Dot- und Short-Anteile weisen also auf Reads hin, die entweder zu kurz sind oder zu schlechte Signale produzieren oder auf beides gleichzeitig. Gegen die Vermutung von Beads, die zu schlechte Signale produzieren, sprechen eigentlich die guten Ergebnisse beim Enrichment der Beads. Wenn an den Beads nur wenig DNA amplifiziert worden wäre, dann hätten die Ausbeuten im Enrichment geringer sein sollen; Beads mit wenigen Amplifikaten sollten die Waschschrte beim Beadrecovery im MPC nicht überstehen.

Für die Vermutung, dass die Reads insgesamt kurz sind (und damit die Amplifikationsprodukte an den Beads), spricht die große Zahl an Reads, die als „Failed Short Primer“ heraussortiert wurden; anscheinend wurde bei einer Vielzahl von Reads 454 Adapter-Sequenzen am Anfang erkannt (Adapter-B-Sequenz). Es scheint also Probleme mit der Amplicon Library-DNA selber zu geben.

Schaut man sich die Gelbilder der Amplicon Librarys an bzw. die Größenverteilung der Positionen 6, so liegt der Hauptteil zwischen 0,5 und 1 kb, und es ist keine Ansammlung von kleineren Fragmenten zu erkennen. Dazu kommt noch, dass diese Librarys danach zweimal SPRI-aufgereinigt wurden bei einem Größencutoff von 500 bp; das Auftreten von kleinen Fragmenten ist demnach mehr als unwahrscheinlich.

Um die Widersprüche zwischen den guten Enrichment-Rates und den Gelbildern der Librarys auf der einen Seite und dem Run-Ergebnis auf der anderen Seite aufzulösen bzw. zu versuchen, eine Erklärung für den katastrophalen Run zu finden, sind Tests an der Library-DNA nötig. Die Fehlersuche wurde mittels Amplifikations-Tests an den aqf-Amplicon Librarys durchgeführt und ist im folgenden Abschnitt beschrieben. Die bestehende aqf-Amplicon Librarys können jedenfalls nicht für weitere Runstarts verwendet werden.

4.2.2.5 Amplifikations-Test mit den aqf-Amplicon Librarys

In einem ersten Versuch wurde erprobt, ob die aqf-Amplicon Librarys funktionsfähige Adapter-Sequenzen tragen. Dazu wurden diese in einer PCR mit Primern, die die 454-Adapter-A- und -B-Sequenzen darstellen, amplifiziert. Es wurde getestet, in welchem Größenbereich die Amplifikationsprodukte der aqf-Amplicon Librarys liegen. Wenn die Amplicon Librarys funktionieren würden, müssten die Produkte einer PCR auf diesen Librarys mit den Adapter-A- und -B-Sequenzen als Primer idealerweise im selben Größenbereich wie die Librarys liegen. Wie das zugehörige Gelbild zeigt, ist dies aber nicht der Fall. Bei allen Librarys sind die Amplifikationsprodukte sehr klein, liegen weit unterhalb von 500 bp. Die Größenverteilung der Amplifikationsprodukte entspricht dem des Sequencing-Run-Ergebnisses; der Hauptteil ist sehr klein und in geringen Mengen sind auch Produkte bis hin zu 500 bp vorhanden (der DNA-Schmier zieht sich bis dorthin bei den meisten Librarys). Auch bei dem Sequencing-Run der Amplicon Librarys gab es einige Reads, die zumindest größer als 300 bp waren –im Signalprocessing werden die Reads durch die verschiedenen Filter gekürzt, so dass die Passed- Filter Wells selten so lang bleiben wie die ursprünglich Sequenz.

Das Ergebnis des Amplifikations-Tests bestätigt, dass die Librarys das Problem sind und wirklich so kurze Sequenzen liefern, d.h. die kurzen Reads im Sequencing-Run kamen hauptsächlich nicht aufgrund von möglichen schwachen Signalstärken zustande, sondern aufgrund von Library-DNA, die tatsächlich nur aus kurzen DNA-Sequenzen zu bestehen scheint. Die dicken Amplifikationsprodukte, die auf dem Gelbild des Amplifikations-Tests zu sehen sind, lassen sogar vermuten, dass die Signalstärken nicht zu niedrig gewesen sind. Die kurzen Amplifikationsprodukte erklären dann auch den hohen Anteil an Reads, die als „Failed Short Primer“ aussortiert wurden. Die Ergebnisse des Amplifikations-Tests, die die Sequencing-Ergebnisse bestätigen bzw. in die Richtung verfestigen, dass amplifizierbare Library-DNA nur in kurzer Sequenzlänge vorlag, deuten auf zwei Probleme hin. Auf der einen Seite scheinen die PCR-Produkte, die bei der Amplicon Library-Herstellung den Hauptteil ausmachen in einem Größenbereich von 0,5 bis 1 kb Länge keine funktionsfähigen Librarys darzustellen. Es sind „echte“ Amplifikationsprodukte, die bei der Generation der Amplicon Library erstellt wurden, sie liegen auch in dem richtigen Größenbereich vor, scheinen aber entweder fehlerhafte oder keine 454-Sequenzen zu tragen. Mögliche Ursachen könnten Sequenz-Ähnlichkeiten von 454-Adapter-A- und -B-Sequenzen und den MINT-Adaptersequenzen sein, oder

Wechselwirkungen mit der Template-DNA, da die verwendeten Primer bei der Amplicon-Herstellung sehr lange 5'-Overhangs besitzen (der Forward-Primer besitzt ein 40 bp langes 5'-Overhang, der Reverse-Primer ein 30 bp langes). Auf der anderen Seite scheint die doppelte SPRI-Aufreinigung nicht die kurze Library-DNA in ausreichendem Maß reduziert zu haben. Das ist mit Ausnahme von Handling-Fehlern nicht erklärbar, da die Ampurebead-Charge bei nachfolgenden Aufreinigungen von genomischen Shotgun-Librarys funktioniert hat (Ergebnisse werden nicht gezeigt). Aber selbst wenn die SPRI-Aufreinigungen versagt hätten, dann hätte es zumindest auch größere Amplifikationsprodukte beim Amplifikations-Test geben müssen, wenn die Library-DNA selber in Ordnung wäre.

Da mit diesen Amplicon Librarys nicht weitergearbeitet werden konnte, sollte versucht werden, mögliche Fehlerquellen zu identifizieren bzw. auszuschließen, um durch Veränderungen im Protokoll doch noch eine 454-Library-Produktion zu ermöglichen. Im folgenden Abschnitt wird beschrieben, ob das Multiplex-Template ein Problem darstellt –mögliche kleine Fragmente könnten sich in der PCR während der Amplicon Library-Herstellung bei der Amplifikation durchgesetzt haben und somit Ursache für die kurzen Sequenzen sein.

4.2.2.6 Amplifikations-Tests mit aqf-Sanger-Klonen

Um ausschließen zu können, dass die Probleme, die bei dem Sequenzieren der Amplicon Librarys auftraten, durch das Multiplex-Template in der PCR bei der Herstellung der Librarys verursacht wurden, wurde die selbe Prozedur zur Herstellung von Amplicon Librarys mit den gleichen Primern und PCR-Mix auf Sanger-Klone angewendet. Aus fünf Sanger-Klonen der mittels MINT-Kit erstellten cDNA wurde in einer Gradienten PCR versucht, Amplicon Librarys herzustellen. Das Gelbild dieser PCR zeigt, dass ein entsprechendes Amplifikations-Produkt bei allen fünf Sanger-Klonen über den gesamten Annealing-Temperatur-Bereich erzeugt wurde. Die Amplifikations-Produkte sind um 1 kb groß (die Größen der verschiedenen Sanger-Klone variieren leicht, so ist beispielsweise aqf3b_4 etwas kleiner als aqf3b_8) und es sind dementsprechend Banden. Es zeigt sich also ein Bild, wie man es von einer vermeintlich erfolgreichen Amplicon Library-Herstellung erwartet, da die Produkte die richtigen Größen und Form (Banden) besitzen. Die dazugehörige Negativkontrolle ist vermutlich nicht ganz frei von Produkt, da leichte Banden bei ca. 80 bp zu sehen sind. Es ist nicht auszuschließen, dass es sich um ein „Nonsens“-Amplifikat handelt, da die in der PCR verwendeten Primer mit 56 bp (ADA-MID1-CDNA) bzw. 49 bp Länge (ADB-CDNA) eigentlich etwas zu kurz sind, um diese leichten Banden in der Größe zu erklären. Da in den Ansätzen mit den Sanger-Klonen keine Banden oder Bandenschatten in dem Bereich von 80 bp auftreten, tritt das mögliche Nonsens-Amplifikat vermutlich nur dann auf, wenn kein spezifisches Template für die Primer vorhanden ist.

Die Amplifikationsprodukte der Gradienten-PCR wurden nun für eine Test-PCR mit den 454-Sequenzen als Primer verwendet. Und auf dem Gelbild dieser PCR zeigt sich das gleiche Muster wie schon bei den aqf-Amplicon Librarys: bei den vermeintlichen Amplicon Librarys der aqf-Sanger-Klone ist kein Produkt in zu erwartender Größe zu sehen. Es hätten die gleichen Bandengrößen in allen Ansätzen wie auf dem Gelbild der Gradienten-PCR zu sehen sein müssen, wenn die Amplicon Librarys der Sanger-Klone funktionieren würden. Stattdessen gibt es keine Produkte –außer bei den Ansätzen C1 bis C5, dort sind dünne Banden weit unterhalb von 500 bp zu erkennen (vermutlich

handelt es sich dabei um die gleichen „Produkte“ wie in der Negativkontrolle der vorhergegangenen PCR). Es handelt sich also um die gleiche Problematik wie bei der Herstellung der Amplicon Librarys, mit denen der Sequencing-Run gestartet wurde: eine erste Amplifikation, also die Erstellung der Amplicon Library zeigt Produkte in den erwarteten Größenbereichen, und eine Testamplifikation auf diesen Produkten mit den 454-Adapter-Sequenzen als Primer fällt negativ aus, man erhält kein Produkt bzw. kein Produkt in der erwarteten Größe. Dass die Testamplifikation mit 454-Adapter-Sequenzen prinzipiell funktioniert, zeigt die Positiv-Kontrolle; hier entsteht ein PCR-Produkt in der erwarteten Größe und Verteilung.

An dem Versagen bei der Herstellung der Amplicon Librarys scheint also nicht der Multiplex-Templatecharakter der aqf-cDNAs schuld gewesen zu sein, da auch bei Sanger-Klonen keine Amplicon Library-Herstellung gelingt.

In einem zweiten Anlauf wurde versucht, aus einem Sangerklon mittels Two-Step-PCR eine Ampliconlibrary zu erstellen. Der erste Schritt der Two-Step-PCR stellte dabei die Amplifikation des Sangerklons mit kurzen PCR-Primern dar, die an die durch das MINT-Kit eingeführten Sequenzen der cDNA binden. Der erste Amplifikationsschritt scheint erfolgreich verlaufen zu sein, da Banden in der richtigen Größe vorhanden sind. Es wurden also Amplifikate erzeugt, die von den Primersequenzen umrahmt sind. In einer zweiten PCR sollte an diese Produkte die 454-Sequenzen heranamplifiziert werden mit den langen Primern, die schon bei den ersten Amplicon Library-Versuchen verwendet wurden. Die Hoffnung bestand darin, dass mögliche Fehlpaarungen oder Probleme, die durch den langen 5'-Overhang entstehen, vielleicht verringert oder beseitigt werden könnten, wenn man ein Template in der PCR verwendet, das besser an diese angepasst ist. Da durch den ersten Amplifikationsschritt Template-DNA gewonnen wurde, deren 5'-Enden den 3'-Enden der Primer der zweiten PCR entsprechen und nicht darüber hinausgehen, könnten mögliche Probleme verursacht durch die langen 5'-Overhangs möglicherweise verringert werden. Das Gelbild der zweiten PCR zeigt, dass diese Amplifikation fehlgeschlagen ist. Das PCR-Mix bzw. die PCR selber hat funktioniert, die verwendete Positivkontrolle zeigt Produkte in der richtigen Größe und die Negativkontrolle ist frei von Produkten. Im Gegensatz zu der Herstellung der aqf-Amplicon Librarys ist bei dieser PCR also noch nicht einmal ein Produkt zu sehen, obwohl das Template eigentlich geeigneter sein sollte: es ist ein Sangerklon, hat also keinen Multiplexcharakter, und es ist durch den ersten Schritt der Two-Step-PCR zurechtgeschnitten, d.h. es gibt beim Template selbst keine störenden und nichtbindenden 5'-Overhangs. Dies ist ein weiterer Hinweis darauf, dass die Primer mit den langen 5'-Overhangs, die die 454-Sequenzen tragen, ein Problem darstellen, sei es aufgrund ihres Overhangs oder der Sequenz, die Ähnlichkeiten zu den MINT-Sequenzen aufweist.

Da es fraglich ist, ob durch die Wahl anderer Primersequenzen eine Amplicon Library-Herstellung aus den aqf-cDNAs mittels Two-Step-PCR funktioniert –denkbar wären auch vielleicht mehr als nur zwei PCRs, bei denen nur kurze 5'-Overhangs verwendet werden (etwa 5 bp) und man sich so bis zu den vollständigen 454-Sequenzen in fünf bis sechs PCRs vortastet- und mit jeder weiteren PCR die Qualität des Templates leidet (Vergrößerung des Bias) und der Aufwand steigt, musste eine andere Methode gefunden werden, um die aqf-cDNAs auf dem 454 erfolgreich zu sequenzieren.

4.3 Entwicklung einer neuen Methode zur gerichteten Sequenzierung

Es sollte eine neue Methode gefunden werden, mit der es möglich ist, einerseits aus den vorhandenen aqf-cDNAs sequenzierbare 454 Librarys herzustellen, andererseits soll diese Methode idealerweise auch auf andere DNA-Templates, die nicht die MINT-spezifischen Sequenzen enthalten, anwendbar sein. Dazu muss sie eine gerichtete Sequenzierung ermöglichen, als Alternative zur Erstellung von Amplicon Librarys.

Den Kern dieser neuen Methode bildet eine gerichtete Ligation. Die zur 454-Pyrosequenzierung benötigten Adapter-A- und -B-Sequenzen werden an die Template-DNA heranligiert. Im Unterschied zu den bereits vorhandenen und etablierten Methoden werden Adapter benutzt, die an der Template-DNA zugewandten Seite Sticky-Ends besitzen. Über unterschiedliche Sequenzen der Sticky-Ends von Forward- und Reverse-Adapter wird eine gerichtete Ligation ermöglicht. Zur Ligation wird dabei T4-DNA-Ligase verwendet [126], die auch kurze, 3 bp lange Überhänge effizient zusammenführen kann [127].

4.3.1 Erstellung einer Test-Library

In einem ersten Test wurde versucht, aus der aqf65-cDNA eine 454-Library herzustellen. Es wurde eine relativ große Menge an DNA von 5 µg verwendet, da nicht abgesehen werden konnte, wie effizient die Library-Herstellung funktionieren würde und wie hoch die Verluste während der unterschiedlichen Aufreinigungsschritte wären. Außerdem kann dann ein entsprechender Vergleich mit dem General-Library-Protokoll stattfinden, da hier die gleiche Ausgangs-DNA-Menge zum Einsatz kommt.

Die cDNA wird mit dem Restriktionsenzym SfiI geschnitten, da durch das MINT Kit entsprechende Erkennungssequenzen eingeführt wurden. SfiI ist eine Typ II Restriktionsendonuklease aus *Streptomyces fimbriatus*, deren Aktivität bei einer Temperatur von 50°C am größten ist [128], weshalb der Restriktionsverdau bei 50°C stattfand. Bei dieser Temperatur sollte auch ein Reannealing als eine Art Konkurrenzreaktion zum Verdau unterbunden werden. SfiI liegt als Tetramer vor, das an zwei Erkennungssequenzen binden muss, bevor es schneiden kann [129,130]. Dabei werden DNA-Substrate mit zwei oder mehr Erkennungssequenzen bevorzugt, die weiter als 150 bp auseinanderliegen [131-133] –die aqf-cDNA sollte damit ein ideales Template sein.

Für den erfolgreichen Fortgang der Library-Herstellung nach der Restriktion muss sichergestellt werden, dass das Restriktionsenzym selber inaktiviert wird, und dass die abgeschnittenen Termini der MINT-Konstrukte entfernt werden, bevor die Ligation startet. Beides sollte mit der Aufreinigung mit dem MinElute-Kit erreicht werden: die verwendeten Säulen sollen laut Angaben des Herstellers keine DNA-Fragmente kleiner als 40 bp binden, womit die abgeschnittenen Termini entfernt würden; außerdem enthält der bei der Aufreinigung verwendete PB-Puffer Guanidinhydrochlorid, das Proteine denaturiert (womit SfiI funktionsunfähig gemacht werden müsste).

Die anschließende Ligation mit T4-DNA-Ligase findet mit selbst hergestellten Adaptern statt. Diese besitzen nur an dem 5'-Ende der überhängenden Seite ein Phosphat, wodurch einerseits eine Ligation zur Insert-DNA ermöglicht werden soll (die nach der Restriktion ihrerseits an den 5'-Ende Phosphatreste besitzt), und andererseits verhindert wird, dass es zu Blunt-End-Ligationen der Adaptermoleküle kommt. Ferner soll durch die Temperatur von 4°C eine Sticky-End-Ligation bevorzugt werden (Angaben des Herstellers). Die verwendete Adaptermenge entspricht dem bei der General-Library-Prozedur angewandten Verhältnis von 15:1 (Adaptermenge zu eingesetzter DNA), da damit durchgehend positive Erfahrungen gemacht wurden –die Ausbeuten an funktionsfähiger Library-DNA waren gut (d.h. messbar auf einem RNA-Pico-Chip) und auch die dort eingesetzte Adaptermenge nicht zu groß, so dass nach den SPRI-Aufreinigungen noch Adapterreste enthalten waren.

Nach Hitzeinaktivierung der Ligase findet erneut eine Aufreinigung des Ansatzes mit dem MinElute-Kit statt. Dies geschieht um sicherzustellen, dass die Library-DNA in hoher Qualität vorliegt –außerdem erhöht es ebenfalls die Vergleichbarkeit mit der General-Library-Prozedur, da hier nach jedem enzymatischen Schritt ebendiese Aufreinigung erfolgt. Die beiden aufeinanderfolgenden SPRI-Aufreinigungen sollen sicherstellen, dass einerseits die Library-DNA nicht unterhalb von 500 bp vorliegt, andererseits sollen unverbrauchte Adapter entfernt werden.

Die neue Prozedur beinhaltet mehrere Faktoren, deren Gelingen notwendig ist aber erst im Ergebnis der Test-PCR ausgewertet werden können: dass Restriktion und Entfernen von SfiI erfolgreich waren, die Herstellung der Adapter mit den Sticky Ends funktioniert und diese sich an das geschnittene Template ligieren lassen und dass die Ligationsprodukte von kleineren Fragmenten (auch Adapter) aufgereinigt werden.

4.3.1.1 Test-PCR

Anhand des Gelbildes der Test-PCR ist ersichtlich, dass die Testlibrary erfolgreich amplifiziert werden konnte. Da die verwendeten Primer aus Oligonukleotiden mit den Adapter-A- und -B-Sequenzen bestanden, muss die Template-DNA, also die Test-Library, die entsprechenden Sequenzen besitzen. Das bedeutet, dass die vorangegangene Prozedur erfolgreich gewesen sein muss, die Sticky-End-Ligation der Adapter an eine vorher mit dem Enzym SfiI geschnittene aqf65-cDNA muss funktioniert haben. Die Amplifikationsprodukte der Test-Library liegen auch in einem erwarteten Größenbereich von 0,5 bis 2 kb vor, d.h. die Erstellung der Test-Library war auch über den gesamten Größenbereich erfolgreich. Ferner kann man auf dem Gelbild erkennen, dass es so gut wie kein Produkt unterhalb von 500 bp gibt, d.h. durch die SPRI-Aufreinigung wurden mögliche kleinere Fragmente entfernt.

Im Gegensatz zur Amplicon Library-Methodik liefert die Test-PCR hier also ein anderes Bild: die Test-Library zeigt keine negativen Auffälligkeiten bzw. der Test fällt positiv aus, es spricht zu diesem Zeitpunkt nichts gegen ein Sequenzieren auf dem 454. Anhand letzterem kann auch nur eine endgültige Aussage getroffen werden, ob und wie gut die neue Methode zur Herstellung von gerichteten 454-Librarys funktioniert. Mögliche Fehlpaarungen bei der Ligation, Dimer- und Chimären-Bildung können anhand der Test-PCR nicht ausgeschlossen werden.

4.3.1.2 emPCR und Beadrecovery der Test-Library

Da letztendlich trotz positivem Test-PCR-Ergebnis der Test-Library nicht vorausgesagt werden kann, ob diese im Sequencing funktioniert, sollten zuerst kleine Mengen sequenziert werden. Es sollte demzufolge auch nur eine kleine Menge in der emPCR angesetzt werden. Um schnelle Ergebnisse zu bekommen, und da Titrationsen nicht immer aussagekräftig zum Bestimmen des cpb-Wertes sind, wie sich u.a. bei den aqf-General- und –Amplicon Librarys zeigte, wurde beschlossen, die aqf-Test-Library gleich in einer emPCR unterzumischen anstatt sie zu titrieren.

Die Idee dabei war, als „Leitprobe“ in der emPCR eine Library zu verwenden, die schon öfter erfolgreich sequenziert wurde und deren cpb-Wert entsprechend bekannt und eingestellt ist –in dieser emPCR fungiert die Library „271-15“ als Leitprobe. Sie belegt in der Berechnung mit 47,1% den Großteil der Beads (das entspricht 94,2% der Beads des Capture-Bead-Cups, in dem die Librarys zusammen pipettiert wurden); die Test-Library, deren cpb-Wert einfach auf Eins gesetzt wurde, belegt mit 1,5% (oder 3% des Cups) nur einen sehr geringen Teil. Sollte der cpb-Wert von Eins zu klein sein, so dürften einfach nur zu wenig Beads (und damit Sequenzen) der Test-Library entstehen, was aufgrund des geringen Anteils an den Beads nicht ins Gewicht fällt für die Möglichkeit eines Run-Startes. Sollte der cpb-Wert zu groß sein, so entstünden zwar mehr Beads als erwartet, da aber die Test-Library auf einen sehr geringen Anteil im Vergleich zur Leitprobe berechnet wurde, sollten die Auswirkungen nicht zu groß werden, d.h. die Enrichment-Rate sollte nicht sehr stark ansteigen und die Beads immer noch sequenzierbar sein. Nach einem erfolgreichen Sequencing-Run würde man anhand des Anteils der Test-Library an den erzeugten Reads erkennen können, wie stark man sich mit dem cpb-Wert verschätzt hat und diesen versuchen entsprechend anzupassen für weitere Runs. Im Vordergrund beim Untermischen bzw. Sequenzieren der Test-Library steht jedoch nicht die Bestimmung des idealen cpb-Wertes, sondern die Sequenzierbarkeit der Library an sich, ihre durchschnittliche Größe und Verteilung.

Das nach der emPCR stattgefundenene Beadrecovery zeigt mit 8,8% eine gute Enrichment-Rate. Mit knapp 2,6 Mio. Beads wurden auch etwas mehr erhalten, als für die Beladung einer Region nötig sind (ca. 2 Mio.); mit diesen Beads kann also ein Laufstart durchgeführt werden.

4.3.1.3 Sequencing-Run mit der Test-Library

Die aufgereinigten Beads, die auch die Test-Library enthalten, wurden auf eine Region einer PTP geladen und sequenziert. Mit über 550000 Reads bei einer durchschnittlichen Leselänge von 365 bp auf Region 1 liefert dieser Sequencing-Run ein gutes Ergebnis. Die herausgefilterten Reads zeigen auch keine Auffälligkeiten, wenn auch der Anteil von 27% an „Failed Short“ etwas geringer sein könnte –bei optimalen Runs liegt er knapp unter 20%. Die Kontroll-DNA-Reads zeigen keine Besonderheiten, ihre Ausbeuten sind so hoch (Passed Filter Wells und Anteile über alle Längen bei den unterschiedlichen Übereinstimmungen), wie man es für einen guten Run erwarten würde. Das Ergebnis des Runs folgt auch den Erwartungen an die erhaltene Enrichment-Rate der Beads von 8,8% –die Ausbeute an erhaltenen Reads von der Region liegt innerhalb der Vorgaben und es wurden nur wenige Reads als Mixed entfernt. Der Anteil nicht monoklonaler Beads ist also gering, was bei dieser Enrichment-Rate von 8,8% auch zu erwarten war.

Von den 552883 Reads der Region 1 entfallen 19749 auf die Test-Library, das entspricht einem Anteil von 3,57%. Die Test-Library war auf 1,5% der gesamten Beadmenge eines emPCR-Kits berechnet. Da sie mit den anderen Librarys der Region 1 in einem Capture-Bead-Tube angesetzt worden war - ein emPCR-Kit besteht aus zwei Capture-Bead-Tubes, wobei jedes Tube je 50% der gesamten Capture-Bead-Menge enthält - hätte die Test-Library laut Berechnung genau 3% der Beads ausmachen müssen. Die Differenz zu dem tatsächlichen Anteil an Reads ist minimal, mit anderen Worten, der verwendete cpb-Wert von Eins war ideal, wenn auch ein Zufallstreffer.

Entscheidend zur Beurteilung der Qualität der Test-Library ist deren durchschnittliche Leselänge und Größenverteilung. Mit einer durchschnittlichen Leselänge von 355 bp und der größten Fraktion zwischen 401 bis 500 bp bei 39,2% der knapp 20000 Reads ist die Test-Library ein Erfolg. Es ist damit nicht nur gelungen, zum ersten Mal eine sequenzierbare 454-Library aus der aqf-cDNA herzustellen, sondern auch die Qualität der Library ist gut. Das bedeutet, dass die neue Methodik der Library-Herstellung über gerichtete Ligation einwandfrei funktioniert hat und diese es ermöglicht, cDNAs gerichtet zu sequenzieren, bei denen eine Erstellung einer Amplicon Library gescheitert ist.

4.3.2 Sequenzieren der aqf-Librarys mit der neuen Methodik

Da die Ergebnisse der Test-Library positiv ausgefallen waren und die neue Methodik zu funktionieren schien, sollte sie dementsprechend bei allen aqf-cDNA angewendet werden, damit sie alleine auf einer PTP-Region ohne andere Librarys sequenziert werden konnten. Zum einen wurden selbstverständlich die Sequenzen gebraucht, zum anderen lässt sich nur so abschließend bestätigen, ob die neue Methodik funktioniert und generell geeignet ist für das 454-Sequencing. Die Vortests fielen alle positiv aus, aber bis zu diesem Zeitpunkt konnte noch nicht ausgeschlossen werden, dass es nicht doch zu bis jetzt unbekannten negativen Effekten und Auswirkungen kommt, die sich erst in der Masse, also bei voller Beladung einer Region mit den neuen Librarys, zeigen (beispielsweise ähnlich zu den Überstrahlungen durch die Poly-T-Bereiche bei dem General-Library-Sequencingrun).

4.3.2.1 Erstellen der Librarys

Anhand der neuen Methodik wurden aus den zwölf aqf-cDNAs 454-Librarys hergestellt. Wie bei der Test-Library fand auch hierbei anschließend eine Test-PCR mit den Adapter-A- und -B-Sequenzen als Primer statt. Auch hierbei zeigten die erstellten 454-Librarys Amplifikationsprodukte im erwarteten Größenbereich. Dementsprechend müssen die hergestellten Librarys die Adapter-A- und -B-Sequenzen besitzen, die über die gerichtete Ligation an die Insert-DNA gebracht wurde. Zudem liegt der Schwerpunkt der amplifizierten DNA im Bereich von über 0,5 kb, d.h. auch die SPRI-Aufreinigung war erfolgreich.

Die Ergebnisse der Test-PCR bestätigen demnach, dass die Erstellung von 454-Librarys nach der neuen Methodik aus den 12 aqf-cDNAs ebenfalls funktioniert hat und sie zum Sequencing angesetzt werden können.

4.3.2.2 emPCR und Beadrecovery der ersten sechs aqf-Librarys

Für einen ersten Sequencing-Runstart wurden sechs der zwölf aqf-Librarys in einer emPCR angesetzt und sollten dann auf einer Region sequenziert werden. Es wurden nicht gleich alle zwölf aqf-Librarys auf einmal angesetzt, da auf der zweiten PTP-Region Standard-General-Librarys sequenziert werden sollten. Sollte die Region mit den aqf-Librarys ein Fehlschlag sein (wie die Runs mit den aqf-General- und –Amplicon Librarys), so könnten die Ergebnisse der zweiten Region Hinweise für die Ursache liefern, da die dort verwendeten Proben schon zu guten Run-Ergebnissen in vorangegangenen Sequenzierläufen führten, womit ein mögliches Versagen des gesamten Runs wegen schlechter Library-Qualität anhand der zweiten Region als Ursache ausgeschlossen werden könnte.

Die beim emPCR-Ansatz der aqf-Librarys verwendete cpb-Zahl wurde auf 1,5 geschätzt. Auch hier wurde wieder auf eine Titration der einzelnen Librarys verzichtet, da es immerhin sechs sind, die hätten titriert werden müssen –ein zeitintensiver Vorgang, außerdem lieferten die Titrations wie schon beschrieben nicht immer eindeutig reproduzierbaren Ergebnisse (auf einen großen emPCR-Ansatz bezogen). Die Test-Library lieferte mit einer cpb-Zahl von Eins gute Ergebnisse, d.h. der Anteil an erhaltenen Reads entsprach ziemlich genau dem vorherberechneten Anteil an Beads. Das war ein guter Anhaltspunkt. Die neuen aqf-Librarys bestanden zwar aus anderen cDNAs als Inserts, aber die Methode der Herstellung war identisch mit der der Test-Library, genauso wie die Methode der Herstellung der cDNAs selber, weshalb ein Versuch der Übertragung der Ergebnisse sinnvoll erschien. Die cpb-Zahl wurde etwas erhöht (1,5), um sicherzustellen, dass genug Beads nach dem Recovery vorhanden wären. Das Beadrecovery lieferte dann auch genug Beads für eine Region bei einem Runstart, auch die Enrichment-Rate lag mit 14,1% noch in dem Bereich, der auf gut sequenzierbare Beads hinweist.

4.3.2.3 Sequencing Run mit den ersten sechs aqf-Librarys

Die im Beadrecovery aufgereinigten Beads mit den aqf-Librarys wurden auf PTP-Region 1 geladen und sequenziert; Region 2 bestand wie oben erwähnt aus Beads mit General-Librarys. Die Ergebnisse des Sequencing Runs sind gut. Die Ausbeuten an erzeugten Reads liegen mit knapp über 530000 Reads auf Region 1 und 600000 auf Region über der von Roche vorgegebenen Untergrenze von 400000 Reads pro Region für einen guten Run mit genomische Shotgun-Librarys –für Amplicons liegt diese bei 375000 pro Region; auch die durchschnittlichen Leseweiten sind mit 360 bp auf Region 1 und 410 bp auf Region 2 besser als die Untergrenze von 350 bp für genomische Shotgun-Librarys. Region 1 mit den nach der neuen Methodik hergestellten aqf-Librarys liegt also auch innerhalb der Spezifikationen für genomische Shotgun-Librarys und zeigt bessere Ausbeuten, als sie für Amplicon Library-Runs erwartet werden. Die Filterwerte der Failed-Reads auf Region 1 zeigen keine Auffälligkeiten; die „größte“ Fraktion wird von denen gestellt, die als Short herausortiert werden (gleiches gilt für Region 2), was bei guten Sequenzierläufen von genomischen Shotgun-Librarys zu erwarten ist, ebenso ihr Anteil von 20%.

Einzig ein wenig auffällig auf Region 1 sind bei der Kontroll-DNA die Anteile an den Reads bei 100%iger Übereinstimmung über alle Längen; sie sind etwas niedriger als für einen so guten Sequenzier-Lauf zu erwarten wäre. Besonders über Längen ab 200 bp sind die Anteile auf Region 1 der Kontroll-DNA bei 100%iger Übereinstimmung in etwa nur halb so groß wie erwartet –selbst bei

dem katastrophalen aqf-Amplicon-Run waren diese besser. Die Ergebnisse der Kontroll-DNA bei Übereinstimmungen von 98% und 95% auf Region 1 sind dagegen gut und liegen bei Ausbeuten, die einem guten Run mit genomischen Shotgun-Librarys entsprechen. Der Sequencing-Run selber zeigt gute Ergebnisse, es gibt anhand der Filterwerte keine Hinweise auf negative Beeinflussung der Kontroll-DNA durch die Library-Reads. Dot- und Mixed-Filterwerte der Library-Werte sind sehr klein und es gab keine Überstrahlungen. Die durchschnittliche Länge der Kontroll-DNA auf Region 1 ist mit rund 386 bp unauffällig und gut. Weshalb es also zu den schlechten Ausbeuten der Kontroll-DNA bei 100%iger Übereinstimmung kommt, kann nicht aus den Run-Ergebnissen hergeleitet werden. Vielleicht kam es zufällig bei einigen Flows während des Sequencings auf Region 1 zu Problemen, die zu „falschen“ Reads bei der Kontroll-DNA führten, weil eine oder wenige Basen falsch bestimmt waren, und damit die Ausbeuten bei 100%iger Übereinstimmung verringerten. Eine mögliche Ursache hierfür wären z.B. Luftblasen, die von dem im Gerät vorhandenen Debubbler nicht eliminiert wurden, oder Feststoff-Reste (Salzreste, -Verklumpungen), die in einigen Flows über Region 1 der PTP gespült wurden -Spekulationen, die nicht nachweisbar sind.

Vergleicht man die Ergebnisse der Zerlegung der Reads auf Region 1 nach ihren MIDs mit dem vorherberechneten Ansatz, dann wird deutlich, dass keine Library sehr stark von der Berechnung abweicht. Sie sind alle trotz gewisser Schwankungen gleichmäßig vertreten. Idealerweise hätten alle Librarys zu identischen Anteilen bei den Reads vorliegen müssen. Auf der einen Seite kann es zu Pipettierfehlern beim Ansatz der emPCR kommen, auf der anderen Seite war für jede Library der cpb-Wert auf 1,5 geschätzt worden. Würde man mit diesen Librarys einen neuen Run starten müssen, so würde man in der emPCR angepasste cpb-Werte verwenden, die sich aus den Ergebnissen dieses Runs herleiten: so würde beispielsweise die Library aqf6b, die mit einem Anteil von 9,3% etwas schwächer vertreten war, einen etwas höheren cpb-Wert erhalten (z.B. 2), und aqf10b, die mit 24,8% stärker vertreten war als berechnet, einen etwas niedrigeren (in etwa 1,1). Alle sechs Librarys zeigen auch eine gleiche Größenverteilung, und es ist kein Zusammenhang zwischen dem Anteil an den Reads und der durchschnittlichen Länge der Librarys ersichtlich. Die Herstellung aller sechs gelaufenen aqf-Librarys war demnach erfolgreich und keine ist zu kurz oder besitzt einen hohen Anteil an kurzen Reads. Im Gegenteil, die Größenverteilung der Reads ist so, wie man es auch bei einer genomischen Shotgun-Library erwarten würde: wenig kurze Reads und die größte Fraktion zwischen 400 und 500 bp Länge liegend.

Die Ergebnisse von Region 1 zeigen also, dass es ebenfalls möglich ist, eine ganze Region mit Librarys zu beladen, die nach der neuen Methodik hergestellt wurden. Es kommt zu keinen negativen Effekten, die Sequenzier-Ergebnisse sind gut und liegen auch noch über den Erwartungen an Amplicon Librarys. Dies ist insofern interessant, als dass die neue Methodik als Ausweg erdacht worden war, weil eine Amplicon Library-Herstellung aus den aqf-cDNAs scheiterte bzw. nicht durchführbar war. Mit der neuen Methode können also nicht nur Librarys für eine gerichtete Sequenzierung hergestellt werden, bei denen eine Amplicon-Herstellung scheitert, sondern die Sequenzier-Ergebnisse entsprechen auch eher denen genomischer-Shotgun-Librarys, die nach der General-Library-Methode (bzw. Rapid-Library-Methode) hergestellt worden sind. Damit ist die neue Methodik nicht nur eine Alternative, sondern auch eine Verbesserung gegenüber Amplicon Librarys.

4.3.2.4 emPCR und Beadrecovery der weiteren sechs aqf-Librarys

Für einen weiteren Sequencing-Run wurde nach dem gleichen Prinzip wie bei den ersten sechs aqf-Librarys die emPCR angesetzt: die eine Hälfte der emPCR bestand aus den sechs weiteren Librarys aqf15, aqf25, aqf35, aqf45, aqf55 und aqf65, die andere Hälfte aus General-Librarys. Beim Ansatz der aqf-Librarys wurde ein cpb-Wert von 1,3 verwendet. Dieser orientierte sich an den Ergebnissen des Beadrecovery der ersten sechs aqf-Librarys. Diese waren mit 1,5 cpb angesetzt worden und erbrachten eine Enrichment-Rate von 14,1%. Es sollte ein wenig die Enrichment-Rate verringert werden, um möglicherweise bessere Sequencing-Ergebnisse zu erhalten; auf eine Titration der einzelnen Librarys wurde auch hier wieder aus den zuvor genannten Gründen verzichtet. Wieweit man dabei hätte vorgehen können war nicht vorherzusagen –verkleinert man den cpb-Wert zu stark, kommen möglicherweise nach dem Beadrecovery nicht mehr genug Beads für die Beladung einer PTP heraus. Und weil die anderen aqf-Librarys mit 1,5 cpb gute Ergebnisse im Sequencing lieferten, fiel die Verkleinerung des cpb-Wertes relativ gering aus. Die nach dem Beadrecovery errechnete Enrichment-Rate betrug 13,3%. Damit liegt man zwar unterhalb des Grenzbereiches von 15-20%, die Beads sollten also gute Sequencing-Ergebnisse liefern, und es waren auch mit rund 3,7 Mio. Beads genügend vorhanden für die Beladung einer PTP-Region, aber die Enrichment-Rate ist trotzdem kaum geringer ausgefallen als die der ersten sechs aqf-Librarys. Man hätte die cpb-Zahl also noch stärker verringern können, anhand dieser Ergebnisse vermutlich auf 1 oder 0,9.

4.3.2.5 Sequencing Run mit den weiteren sechs aqf-Librarys

Genauso wie der letzte Run mit den ersten sechs aqf-Librarys stellt dieser Sequencing-Lauf ein gutes Ergebnis dar: Region 1 mit den aqf-Librarys liefert 478000 Library Reads bei einer durchschnittlichen Länge von 364 bp, Region 2 über 660000 Library Reads mit durchschnittlichen Leseweiten von 410 bp. Der Anteil der im Signalprocessing heraussortierten Reads ist auf beiden Regionen dementsprechend klein. Auch die Größenverteilungen der einzelnen Librarys nach der Zerlegung der Reads in ihr MIDs zeigt, dass alle Librarys der Größenverteilung der gesamten Region folgen; kurze Reads zwischen 1 und 100 bp sind nur gering vertreten, die größte Fraktion wird von Sequenzen zwischen 401 und 500 bp gebildet.

Auch wenn 62,27% Passed Filter Wells auf Region 1 einen guten Wert darstellen, so ist dieser doch etwas kleiner als erwartet. Region 1 des vorangegangenen Sequencing-Runs mit den ersten sechs aqf-Librarys hatte 67,2% Passed Filter Wells, bei einer Enrichment-Rate der geladenen Beads von 14,1%. Bei diesem Run hatten die Beads eine Enrichment-Rate von 13,3%. Eine geringere Enrichment-Rate sollte auch auf „bessere“ Beads hinweisen, also auf Beads, bei denen der Anteil von monoklonaler Library-DNA an den Capture-Beads größer ist, was auch zu besseren Sequencing-Ergebnissen führen müsste (aufgrund der eindeutigeren Signalzuordnung). Auf der einen Seite ist der Unterschied von 13,3 zu 14,1% Enrichment-Rate nicht groß, auf der anderen Seite hätten die Beads mit 13% Enrichment-Rate, die bei diesem Run auf Region 1 geladen waren, zumindest nicht zu schlechteren Passed-Filter-Wells führen sollen, die Ausbeuten hätten zumindest nicht kleiner sein dürfen. Allerdings handelt es sich bei den Beads dieses Runs um andere aqf-Librarys als bei dem vorangegangenen; Unterschiede in der Template-cDNA können durchaus zu einem unterschiedlichen Verhältnis von Enrichment-Rate zu Passed-Filter-Wells führen, was eine mögliche Erklärung wäre.

Allerdings konnte in mehreren unterschiedlichen Sequenzier-Läufen in anderen Projekten mit derselben Library beobachtet werden, dass es bei gleichem Library-Aliquot, cpb-Zahl und gleichem emPCR-Ansatz trotzdem zu Variationen bei der Enrichment-Rate und auch bei dem Verhältnis von Enrichment-Rate zu Passed-Filter-Wells kommt; auch innerhalb einer emPCR- und Sequencing-Kit-Charge, so dass diese Effekte hier ebenfalls nicht ausgeschlossen werden kann.

Die Kontroll-DNA zeigt bei diesem Run gute Ergebnisse, die durchschnittlichen Längen sind mit 410 bp auf Region 1 und 393 bp auf Region 2 hoch; ihre Passed Filter-Werte sind mit über 94% auf beiden Regionen sehr gut. Auch die Anteile von Kontroll-DNA-Reads bei allen Übereinstimmungen (100, 98 und 95% Match) über alle Längen sind so groß, wie man es für einen erfolgreichen Run erwarten würde. Auch bei diesem Lauf zeigt sich allerdings bei den 100%igen Übereinstimmungen ein ähnliches Bild wie bei dem vorangegangenen Run mit den ersten sechs aqf-Librarys: besonders über 100 und 200 bp Länge sind die Anteile der Kontroll-DNA-Reads, die eine 100%ige Übereinstimmung besitzen, auf Region 1 geringer als auf Region 2.

Es scheint kein einzelnes Ereignis zu sein, das zufällig auftritt, wie man es nach dem letzten Run hätte annehmen können. Einen möglichen Hinweis für die Ursache liefern die Größenverteilungen der Reads: sowohl bei diesem Run als auch beim vorangegangenen ist auf Region 1 bei 150 bp eine kleine Häufung von Reads zu erkennen, die bei Region 2 bei beiden Runs nicht auftritt. Geht man davon aus, dass es echte kurze Reads sind und keine abgebrochenen, so bedeutet dies, dass sie am 3'-Ende einen Poly-A-Bereich besitzen. Dieser erzeugt ein starkes Signal und könnte zu leichten Überstrahlungen führen, die zwar keine sichtbaren Auswirkungen auf das gesamte Run-Ergebnis haben, aber möglicherweise zu Mismatchen und damit zu verringerten Anteilen der Kontroll-DNA-Reads bei 100%iger Identität führen könnten. Groß kann dieser Effekt, sollte er auftreten, nicht sein, da die 98%igen Identitäten der Kontroll-DNA-Reads bei allen Längen hoch sind. Aber es wäre eine mögliche Erklärung für den prozentualen Abfall des Anteils bei Region 1 der 100%igen Übereinstimmung bei 200 bp Länge. Vermutlich ist der gleiche Effekt die Ursache für den Abfall bei 100 bp Länge (bei 100%iger Übereinstimmung) der Kontroll-DNA-Reads, allerdings kann in diesem Bereich bei den Größenverteilungen der Reads kein Unterschied zwischen Region 1 und Region 2 erkannt werden. Das kann daran liegen, dass zwischen 50 und 100 bp sowieso ein kleines lokales Maximum zu sehen ist, das aufgrund des Signalprocessings zustande kommt, eine mögliche kleine Anhäufung echter kurzer Read-Sequenzen mag dort untergehen und nicht zu sehen sein. Sollten echte kurze Reads die Ursache darstellen, so kann dem entgegengewirkt werden, in dem bei der Library-Herstellung die SPRI-Aufreinigung noch strenger durchgeführt wird, also der Basen-Cutoff etwas nach oben verlegt wird, um mögliche kurze Sequenzen noch besser wegzureinigen.

4.3.3 Erweiterung auf andere Templates

Bisher wurden ausschließlich Libraries aus DNAs erstellt, die SfiI-Schnittstellen erhielten – bei den aqf-cDNAs wurden diese durch das MINT-Kit bei der cDNA-Synthese eingeführt. Die neue Methodik zur 454-Library-Erstellung sollte aber auch flexibel auf andere Template-DNAs anwendbar sein. Daher musste überprüft werden, inwieweit die neue Methodik auf Template-DNAs übertragbar ist, die keine SfiI-Schnittstellen besitzen und bei deren Herstellung das MINT-Kit gar nicht oder nur teilweise zur Anwendung kam.

4.3.3.1 Sequenzierung einer Illumina-Library

In einem ersten Anlauf sollte eine Illumina-Library auf dem 454 sequenziert werden, da zu diesem Zeitpunkt die Kapazitäten für die Illumina-Sequenzierung begrenzt waren. Die OCon-DNA stellte dabei auch ein gutes Template für die obige Fragestellung dar: sie besaß keine SfiI-Schnittstellen, und außerdem sollte die Sequenzierungsrichtung der ursprünglichen Illumina-Library beibehalten werden.

Um eine 454-Library-Erstellung mittels neuer Methodik zu ermöglichen, sollen an die Template-DNA die SfiI-Schnittstellen heranamplifiziert werden. Die Primer bestehen an den 3'-Enden aus Template-spezifischen Sequenzen, das 5'-Ende umfasst die benötigten SfiI-Erkennungssequenzen. Im Gegensatz zur Amplicon Library-Herstellung von Roche sind die 5'-Overhangs der verwendeten Primer mit 16 bp sehr viel kürzer als die der Amplicon-Primer mit 40 bp Länge. Damit verkleinert sich das Risiko von Loopbildungen und Fehlpaarung, die vielleicht die Ursache für das Scheitern der Amplicon Library-Herstellung waren.

Das PCR-Programm hat nur wenige Zyklen, da es zu keiner Produkt-Vervielfachung kommen soll, sondern nur zu einem „Verlängern“ der Template-DNA um die entsprechenden Primersequenzen. Theoretisch sollte ein Zyklus ausreichend sein; da aber nicht vorhergesagt werden kann, wie effizient die PCR arbeitet, wurden fünf Zyklen verwendet. Nach der PCR wurde eine MinElute-Aufreinigung durchgeführt, um die überschüssigen Primer zu entfernen, die Taq-DNA-Polymerasen zu inaktivieren über das dabei verwendete Guanidinhydrochlorid und so eine möglichst qualitativ hochwertige DNA zu erhalten (keine Buffer- und Protein-Reste). Beide Punkte stellen also einen Unterschied zur Amplicon Library-Herstellung dar: die Verwendung von Primern mit deutlich kürzeren 5'-Overhangs und weniger Zyklen in der PCR. Damit erhöhen sich einerseits die Chancen Produkte zu erhalten, die erfolgreich in eine 454-Library umgewandelt werden können, weil die mögliche Fehlerquellen minimiert werden (Primerfehlpaarungen, Fehlamplifikationen), andererseits sollte die Qualität der Library besser sein (als nach Herstellung mittels Amplicon-Protokoll), da weniger Zyklen in der PCR die Gefahr eines möglichen Bias senken.

Nach dem Heranamplifizieren der SfiI-Schnittstellen an die OCon-DNA wurde aus dieser mittels neuer Methodik eine 454-Library hergestellt. Die so hergestellte Library hatte eine Gesamt-DNA-Menge von 940 ng, was gemessen an der Eingangsmenge von 1450 ng einer Ausbeute von rund 65% entspricht. Während der Herstellung unterlief die DNA insgesamt zwei MinElute- und zwei SPRI-Aufreinigungen. Laut den Angaben des Herstellers soll die Ausbeute bei der MinElute-Aufreinigung bei 80% liegen

[134], wobei im praktischen Laboralltag beobachtet wurde, dass sie geringer ist und höchstens bei 50% liegt (Ergebnisse nicht dargestellt). Die SPRI-Aufreinigungen waren so gewählt, dass der Größencutoff bei 200 bp gesetzt ist, analog eines Cutoffs bei 500 bp für größere DNAs. Ein Cutoff bei 200 bp entspricht bei der in Kapitel 3.1 dargestellten Charge 80-85 µl SPRI-Beads auf 100 µl DNA, da hier noch kein Peak bei 100 bp zu sehen ist. Berechnet man für 85 µl die Ratio von 200 bp zu 900 bp, so erhält man einen Wert von 0,58, für die Kontrolle erhält man 0,64. Der Quotient aus der Ratio von 85µl und der Kontrolle stellt eine gute Annäherung an die Ausbeute dar, die man bei der Aufreinigung von Fragmenten mit 250 bp Länge unter Einsatz von 85 µl SPRI-Beads auf 100 µl DNA erhält; diese Ausbeute liegt bei 91% pro SPRI-Aufreinigung. Errechnet man die Ausbeute nach allen Aufreinigungsschritten (je zwei MinElute und zwei SPRI-Aufreinigungen), so sollte diese bei ca. 53% liegen ($0,8 \times 0,8 \times 0,91 \times 0,91$). Dem zugrunde liegen die Annahmen, dass die Angaben des Herstellers bei der MinElute-Aufreinigung stimmen (Ausbeute von 80%), es während aller Aufreinigungs-Schritte keine Verluste durch das Handling gibt (Verbleiben von nicht mitgeführten Resten in Tubes, Säulen, Pipettenspitzen) und vorallendingen die OCon-Template-DNA kaum DNA-Stränge kleiner als 200 bp besitzt, die bei den SPRI-Aufreinigungen verloren gehen würden. Aus dieser sehr positiven Überschlagsrechnung, was die Ausbeute bei den Aufreinigungen angeht, wird deutlich, dass es bei der PCR zum Addieren der SfiI-Schnittstellen auch zu einer Template-Vervielfältigung gekommen sein muss. Da letzteres nicht nötig und unerwünscht ist aus den oben angesprochenen Gründen könnten bei der Generierung neuer 454-Librarys nach dieser Methode weniger Zyklen in der PCR verwendet werden.

Die so hergestellte OCon-Library wurde in einer emPCR mit anderen Librarys eingesetzt und die produzierten Beads auf eine PTP-Region geladen. Gute und unauffällige Ergebnisse der Kontroll-DNA was Längen, Ausbeuten und Übereinstimmungen angeht, und eine Library-DNA-Menge von fast 500000 Reads bei knapp 60% Passed Filter Wells auf der Region zeigen, dass dieser Sequencing-Run erfolgreich war. Mit anderen Worten die „Umgebung“, innerhalb der die OCon-Library sequenziert wurde, zeigt keine Auffälligkeiten und dürfte keine negativen Einflüsse auf die OCon-DNA-Reads haben. Die 111373 OCon-DNA-Reads haben eine durchschnittliche Leseweite von 320 bp (ohne MID). An die OCon-Template DNA wurden durch das Heranamplifizieren der Schnittstellen und die anschließende 454-Library-Herstellung nach der neuen Methodik insgesamt 86 bp zusätzliche Basen angehängt, 48 bp in Forward- und 38 bp in Reverse-Richtung. Subtrahiert man diese angehängten Sequenzen unter Einbeziehung des Sequencingprimers, so kommt man auf eine durchschnittliche Read-Länge von 270 bp. Das entspricht der durchschnittlichen Länge der OCon-Template-DNA (250 bp). Die Ergebnisse des Sequencing-Runs zeigen also nicht nur, dass das Addieren der SfiI-Schnittstellen mittels PCR mit folgender Library-Herstellung nach der neuen Methodik erfolgreich war, sondern auch, dass die OCon-Template-DNA über ihre gesamte Länge sequenziert werden konnte.

Somit können die SfiI-Sites auch spezifisch an Templates heranamplifiziert werden zur 454-Libraryerstellung nach der neuen Methodik, man ist also nicht nur auf die unmittelbaren Enden (MINT-Kit) von Template-DNAs beschränkt. Es sollte also prinzipiell möglich sein, aus jeder beliebigen Template-DNA eine 454-Library zu erstellen. Ein großer Vorteil gegenüber dem Standard-Amplicon-Protokoll ist der, dass bei der Amplifikation zur Addition der SfiI-Sites noch nicht das MID eingebunden

wird. Mit der neuen Methodik kann zeitnah vor Zusammenstellung eines 454-Runs entschieden werden, welche MIDs verwendet werden sollen; bei dem Standardprotokoll findet mit dem Amplifizieren der Target-Region schon das MID-Tagging statt. Die neue Methodik stellt also eine Entkoppelung von PCR und MID-Tagging dar, was eine größere Flexibilität im Laboralltag bedeutet. So können Engpässe bei der Gestaltung von Multiplex-Runs vermieden werden, die aufgrund von MID-Inkompatibilitäten verschiedener Librarys zustande kommen. Das erhöht auf der einen Seite die Effizienz und den Probendurchsatz, auf der anderen Seite wird die MID-Vergabe vereinfacht und Planungszeiträume verkürzt.

4.3.3.2 Antikörpersequenzierung

Für die Arbeitsgruppe Konthur (MPI für molekulare Genetik, Dahlem) sollte ein Deepsequencing von Immunglobulin-cDNAs auf dem 454 durchgeführt werden, da Florian Rubelt aus der Arbeitsgruppe Konthur die Diversitäten von Antikörperrepertoires verschiedener Spender bestimmen wollte.

Mit der Gesamt-RNA peripherer Blutzellen wurde dabei mittels Poly-T-Primer eine Erststrangsynthese durchgeführt, dann fand eine templatespezifische Amplifikation statt, bei der dem Prinzip aus Kapitel 2.2.7.9 folgend die SfiI-Sites in das Template integriert wurden. Nach einem Gelausschnitt des spezifischen Amplifikationsproduktes folgte eine 454-Library-Erstellung nach der neuen Methodik, da eine Herstellung nach Standard-Amplicon-Protokoll scheiterte.

Zum einen ist dies ein weiterer Test für die universelle Anwendbarkeit der neuen Methodik anhand eines speziellen Beispiels, zum anderen ist die Template-DNA ungünstiger für die Heranampifikation der SfiI-Sites als die OCon-DNA des vorangegangenen Kapitels, da zumindest ein in der PCR verwendeter Primer („Target Specific Primer“) innerhalb der Template-DNA liegt und nicht an deren Ende.

Libraryherstellung. Die Größenverteilung einer Library ist auf einem High Sensitivity DNA Chip während verschiedener Stationen des Protokolls gemessen worden. Das Profil des PCR-Produktes nach Erststrangsynthese zeigt zum einen, dass die von Florian Rubelt (AG Konthur) gewählten Primer zu einer spezifischen Amplifikation des Targetbereiches führten, das target-spezifische Produkt (tsp) liegt hochkonzentriert vor und hebt sich deutlich als Peak vor dem Hintergrund ab. Zum anderen sind aber Nebenprodukte vorhanden –vermutlich unspezifische Amplifikationsprodukte und Primerreste, was die Notwendigkeit einer guten Aufreinigung unterstreicht.

Auf dem Konzentrationsprofil nach Gelausschnitt, Restriktionsverdau (mit MinElute-Aufreinigung), Ligation (ebenfalls mit MinElute-Aufreinigung) und einer ersten Ampurebeadaufreinigung kann man erkennen, dass der Peak des tsp länger geworden ist, er liegt jetzt bei etwas über 700 bp. Dies ist mit den hinzuligierten 454-Adapterbereichen zu erklären und stellt gleichzeitig eine Bestätigung für eine erfolgreiche Ligation dar; es ist nur ein Peak zu erkennen und kein zweiter Peak bei 700 bp, der für ein unligiertes Produkt stehen würde. Gleichzeitig ist die Anzahl und die Konzentration der Nebenprodukte verringert worden. Besonders die vorher noch sichtbaren Peaks bei 100 bp und um 300 bp herum sind verschwunden. Der Gelausschnitt und die eine Ampurebeadaufreinigung verliefen also diesbezüglich erfolgreich. Auf der anderen Seite sind immer noch kleine Produkte bei etwas

oberhalb von 35 bp und ein Peak bei kurz über 500 bp erkennbar. Letzterer stellt vermutlich ein Nebenprodukt der PCR dar, erstere könnten auch unverbrauchte Adapterreste aus der Ligation sein; beide unterstreichen jedenfalls die Notwendigkeit einer zweiten Aufreinigung durch Ampurebeads.

In dem Konzentrationsprofil nach der zweiten Ampurebeadaufreinigung ist zu sehen, dass die zuvor noch vorhandenen Verunreinigungen jetzt nicht mehr zu erkennen sind; als einziger Peak ist das tsp zu sehen, was somit gut sequenzierbar sein sollte.

emPCR, Beadrecovery und Sequencing. Die Region, auf der vier Immunglobulin-cDNA-Librarys geladen waren, die nach der neuen Methodik erstellt wurden, erbrachte mit über 410000 Reads bei einer durchschnittlichen Leseweite von 378 bp ein gutes Ergebnis und bestätigt, dass die neue Methodik auch am Beispiel der Antikörpersequenzierung gut funktioniert. Es ist noch anzumerken, dass die Umstellung auf den „GACT“-Key (Rapid-Library-Key) weder bei der Library-Herstellung, noch beim Sequencing zu Problemen führte, wie am guten Runergebnis ersichtlich ist. Da das Rapid-Library-Protokoll für die Herstellung von Shotgun-Librarys aus (meistens) genomischer DNA sehr viel effizienter ist als das ältere General-Library-Protokoll und anstatt 5 µg nur noch 500 ng DNA als Anfangsmenge benötigt, wurden fortlaufend nur noch Rapid-Librarys hergestellt und dementsprechend auch nur noch Librarys nach der neuen Methodik erstellt, die den Rapid-Library-Key „GACT“ trugen, um eine Kompatibilität bei der Runerstellung zu ermöglichen, da auf einer PTP-Region nur ein Key den Librarys zugeordnet werden kann.

Einzig auffällig bei der Region dieses Runs sind die nicht so optimalen Filterwerte. Mit 43% Passed Filter Wells liegt der Run etwas niedriger als z.B. ein Run bestehend aus einer genomischer Shotgun-Library; die Filterwerte entsprechen denen eines Amplicon Library-Runs. Zum einen dürfte dies nicht verwundern, da die Template-DNA ja auch über eine PCR vervielfältigt wurde. Zum anderen lieferten die vorher gezeigten Runs aber schon bessere Filterwerte. Ein Versagen der Sequenzierchemie kann ausgeschlossen werden, da die Kontroll-DNA-Ergebnisse durchweg gut sind. Ob es in der emPCR Probleme gab bzw. ob die Kit-Charge eher suboptimal war, lässt sich zwar nicht ausschließen; allerdings sollten die Filterwerte bei der vorhandenen niedrigen Enrichment-Rate auch bei emPCR-Problemen etwas besser sein.

Es gibt einige Hinweise in den Runergebnissen auf die mögliche Ursache für die schlechteren Filterwerte. Bei der Größenverteilung der Antikörper-Library-Reads sticht eine besonders hohe Anzahl an Reads heraus, die bei ca. 100 bp Länge liegen. Generell bestehen zwei Möglichkeiten, worum es sich dabei handeln könnte: entweder sind das Reads, die tatsächlich nur so lang sind (echte kurze Library-DNA), oder es sind Reads, die von den Trimming-Filtern des Signalprocessing auf eine Länge von 100 bp verkürzt wurden. Gegen letzteres spricht zum einen, dass eine derartige Häufung zum Peak bei ca. 100 bp zumindest ungewöhnlich wäre –es wäre sehr unwahrscheinlich, dass alle Reads mit schlechteren Signalen ausgerechnet auf genau 100 bp von den Trimming-Filtern gekürzt werden. Zum anderen konnte bei der Datendurchsicht eine Vielzahl von 100 bp langen Reads gefunden werden, die alle gleich aussahen: sie besaßen als Insert nur eine Base, ansonsten bestanden sie nur aus PCR-Primersequenzen und am 3'-Ende aus der Adapter-B-Sequenz. Es handelt sich also zumindest zum Großteil um echte kurze Sequenzen. Solche echten kurzen Sequenzen bzw. Library-

DNAs besitzen über ihre Gesamtlänge eine hohe Signalintensität, da sie potentiell in der emPCR an den Capturebeads sehr stark vervielfältigt werden -im Gegensatz zu langer DNA- und könnten die Ursache für die schlechteren Filterwerte darstellen. Möglicherweise führten die hohen Signalintensitäten dieser kurzen DNA-Stränge zu den erhöhten Dot- und Short-Filterwerten.

Schaut man sich die Größenverteilungen der vier Librarys an, die insgesamt gute durchschnittliche Leseweiten besitzen, so fallen zwei Dinge auf: erstens sind die Größenverteilungen bei allen Librarys sehr gut, die größten Anteile werden bei allen Librarys von den Fraktionen der großen Readlängen (>400 bp) gebildet. Zweitens besitzt die Library hpC im Bereich bis 100 bp nur einen minimalen Anteil an Reads; bei den anderen drei Librarys ist dieser größer (bis hin zu 12%). Bei letzteren spiegelt das das Vorhandensein des 100 bp Peaks im Run wider, bei der hpC-Library scheint dieses Produkt nicht aufzutauchen; zumindest ist es in ihrer Größenverteilung nicht erkennbar. Das wiederum deutet darauf hin, dass das Vorhandensein des 100 bp-Fragments, das vermutlich für die schlechteren Filterwerte verantwortlich ist, kein systemübergreifendes Problem bei der Antikörperlibrary-Herstellung darstellt, es tritt nicht generell auf, sondern es ist templatespezifisch. Entweder kam es bei der Herstellung der hpC-Library nie vor oder es war nur so gering vorhanden, dass es fast vollständig weggereinigt werden konnte.

Da dieses 100 bp-Fragment die PCR-Primersequenzen besitzt, scheint es ein Nebenprodukt der PCR zu sein und stellt ein Nonsensamplifikat dar, weil es nur eine Base als Insert enthält. Sein Auftreten im Run unterstreicht nochmal die Notwendigkeit der Aufreinigungsschritte während der Library-Produktion. In den Konzentrationsprofilen der Librarys war zwar kein Peak bei ca. 100 bp nach der zweiten Ampurebead-Aufreinigung mehr sichtbar, er lag unterhalb der Nachweisgrenze. Allerdings schließt das ein Vorhandensein in geringsten Konzentrationen nicht aus; dass es Produkte in diesem Größenbereich gibt, zeigen die Konzentrationsprofile der Librarys nach der PCR. Vielleicht hätte ein dritter Ampurebead-Aufreinigungsschritt die Situation verbessert und zu einem geringeren Anteil des 100 bp-Fragmentes geführt. Die Anzahl der Ampurebead-Aufreinigungsschritte ist aber durch die Template-Menge limitiert; die Ausgangs-RNA lag nicht in beliebig hohen Konzentrationen vor und stellte einen Engpass dar (Rücksprache mit Florian Rubelt).

Zusammenfassend ist zu sagen, dass die neue Methodik zur Libraryherstellung ebenfalls zur Antikörpersequenzierung geeignet ist und gute Run-Ergebnisse liefert; sie funktioniert auch mit „schwierigeren“ Templates, bei denen die Primerbindestellen nicht nur am Ende des Templates liegen, was die universelle Anwendbarkeit unterstreicht. Außerdem wurde sie erfolgreich auf den „GACT“-Key der Rapid-Librarys adaptiert. Je nach Template sollte man aber teilweise einen dritten Ampurebead-Aufreinigungsschritt einplanen, sofern die Ausgangsmengen dies erlauben, um eine hohe 454-Runqualität sicherzustellen.

4.3.3.3 Verbesserungen des Protokolls für die Antikörpersequenzierung

Mit dem vorher verwendeten Protokoll war es möglich geworden, mit Anfangs-Mengen für die Restriktion von bis zu 200 ng auszukommen, um am Ende der Library-Herstellung noch DNA im messbaren Bereich zu erhalten. Manche Proben konnten aber nicht in diesen Mengen zur Verfügung gestellt werden, wie beispielsweise viele Immunglobulin-cDNAs (von Florian Rubelt AG Konthur, MPI für molekulare Genetik Berlin). Es sollte also eine Optimierung des Protokolls stattfinden, um auch aus sehr kleinen Ausgangsmengen 454-Libraries herzustellen.

Die Idee dabei bestand darin, Aufreinigungsschritte auszusparen oder durch andere, weniger verlustreiche zu ersetzen. Die MinElute-Säulenaufreinigung zwischen Restriktion und Ligation zur Inaktivierung des Restriktionsenzym und Entfernung der geschnittenen kleineren Fragmente sollte ersetzt werden (nach eigenen Erfahrungen als auch nach Rücksprache mit Kollegen zeigen MinElute-Aufreinigungen Ausbeuteverluste von über 50%). Da eine Hitzeinaktivierung von SfiI nicht möglich ist (Angaben des Herstellers), soll die benötigte Inaktivierung des Restriktionsenzym mittels Proteinase K stattfinden, einer Protease, die erstmals aus dem Pilz *Tritirachium album* isoliert wurde [135]. Proteinase K kann selber durch Hitze weitestgehend inaktiviert werden (15 min Schritt bei 75°C im Protokoll), eine Restaktivität soll aber weiterhin verbleiben (laut Angaben des Herstellers). Das sollte aber unproblematisch sein, da die folgende Ligation bei 4°C stattfindet und das Aktivitätsmaximum von Proteinase K bei 37°C liegt und mit niedrigeren Temperaturen abnimmt [136], weshalb deren Verwendung zur SfiI-Inaktivierung vertretbar erscheint, da sie die folgende Ligation weitestgehend nicht beeinflussen dürfte. Durch das Verzicht auf die MinElute-Aufreinigung bleiben allerdings die von SfiI abgeschnittenen kleinen Fragmente, die die cDNA flankierten, erhalten. Diese stellen prinzipiell ein Problem dar, da sie in Konkurrenz zur Ligation der Adapter treten. Diesem Problem soll bei dem verwendeten Protokoll doppelt entgegengetreten werden: zum einen werden die Adapter bei der Ligation in einem starken Überschuss zugegeben, zum anderen wird bei der Proteinase K-Inaktivierung nach der Restriktion die Temperatur auf 75°C (für 15 min) erhöht und anschließend auf 10°C heruntergefahren. Die Höhe der Inaktivierungstemperatur müsste bewirken, dass ein Großteil der während der Restriktion abgeschnittenen Fragmente denaturiert vorliegen, da die Schmelztemperatur der Fragmente mit 10 bp langem doppelsträngigen Bereich weit überschritten ist, weshalb diese bei der nachfolgenden Doppelstrangligation keine Rolle spielen sollten. Gleichgeblieben ist die doppelte SPRI-Bead-Aufreinigung nach der Ligation, da diese einerseits verlustarm ist und andererseits unerlässlich zum Entfernen kleinerer unerwünschter Produkte (kleine PCR-Produkte, Adapter).

Ein weiteres Mittel zur Senkung der Anfangs-DNA-Mengen ist das Ersetzen von Größenselektion der cDNA vor der 454-Library-Herstellung. Um auf Gel-Ausschnitte oder sonstige Fraktionierungen vor der Restriktion verzichten zu können, werden die Größenselektionen während der 454-Library-Herstellung durchgeführt. Kleinere unerwünschte Produkte werden durch die doppelten SPRI-Bead-Aufreinigungen nach der Ligation entfernt. Die Aufreinigung vor größeren unerwünschten Fragmenten, beispielsweise Nebenprodukte einer PCR, findet in einem optionalen Größenunterschuss statt. Dieser basiert auch auf der Verwendung von SPRI-Beads, nur dass diesmal die unerwünschten größeren Fragmente an die Beads gebunden werden und die kleineren in Lösung verbleiben, welche nachher in

einem weiteren Schritt an SPRI beads gebunden und so wiedergewonnen werden. Durch diesen optionalen Schritt am Ende des Protokolls soll es einerseits ermöglicht werden, größere Produkte zu entfernen, aber andererseits gleichzeitig die Verluste der 454-Library-DNA so gering wie möglich zu halten. Durch die Auftrennung der Separation kleiner und größerer Nebenprodukte sollten die Ausbeuteverluste ebenfalls gesenkt werden, da Proben, die keine großen Nebenprodukte enthalten, auch nicht diesen Aufreinigungsschritt und damit verbundenem Ausbeuteverlust durchlaufen –im Gegensatz zu z.B. einem Gel-Ausschnitt oder einer Säulenfraktionierung.

Libraryherstellung. Anhand der bestimmten Eingangsmengen an verwendeter cDNA und der erhaltenen DNA-Mengen am Ende ist ersichtlich, dass Eingangsmengen von nur 6,66 ng (RA-BS_03LK) ausgereicht haben, um messbare Library-DNA-Konzentrationen und Mengen zu erhalten; d.h. die 454-Library-Herstellung ist auch mit sehr geringen DNA-Mengen möglich –zur Rapid-Library-Herstellung werden beispielsweise 500 ng benötigt. Die Ausbeuteverluste variieren stark, bei RA-BS_12H bleibt rund ein Viertel der Eingangs-DNA-Menge am Ende erhalten, bei RA-BS_10LK dagegen nur ein Hundertstel. Das liegt vorallendingen daran, dass zu Beginn die Gesamt-DNA-Menge mittels Messung auf dem Qubit bestimmt wurde, es werden also auch kleinere Nebenprodukte erfasst, vielleicht auch Primer-Reste aus der PCR, in der die cDNAs amplifiziert wurden. Kleinere DNA-Fragmente werden durch die zweifache SPRI-Aufreinigung entfernt, so dass die Ausbeuteverluste größtenteils aufgrund der unterschiedlichen Mengen entfernter Fragmente zustande kommen.

Die Library RA-BS_10LK besitzt die geringste DNA-Menge von allen Librarys am Ende der Herstellung, und im Konzentrationsprofil ist nur ein schwacher Peak bei ca. 700 bp zu erkennen. Auf der einen Seite liegt der Peak damit im erwarteten Längenbereich des spezifischen Amplifikationsproduktes, auf der anderen Seite macht die geringe Menge eine Einschätzung über etwaige Nebenprodukte oder Verunreinigungen mit kleinen Fragmenten schwierig (z.B. Adapterreste). Da die Produktion von mehr DNA für eine 454-Library-Herstellung scheiterte, sollte sie trotzdem für einen Sequenzierrun verwendet werden. Die KaretikLK-Library dagegen zeigt einen deutlichen Peak von erwarteter Länge des spezifischen Amplifikationsproduktes bei 600 bp. Man kann anhand dieses Konzentrationsprofils eine bessere Aussage über das Vorhandensein von Nebenprodukten treffen. Kleinere Nebenprodukte sind keine enthalten bzw. liegen unterhalb der Nachweisgrenze. Die doppelte SPRI-Bead-Aufreinigung war also erfolgreich. Auf der anderen Seite sieht man einen zweiten deutlichen Peak knapp unterhalb von 2000 bp. Da dieser nicht das spezifische Amplifikationsprodukt darstellt -diese sollen bedingt durch die Primerwahl zwischen 600 und 850 bp liegen (Rücksprache mit Florian Rubelt, AG Konthur, MPI für molekulare Genetik, Berlin), muss er entfernt werden. Dafür wurde mit der Library der optionale Größenunterschluss durchgeführt, bei dem ebenfalls SPRI-Beads verwendet werden, allerdings im umgekehrten Prinzip: Fragmente, die größer als das spezifische Amplifikationsprodukt sind und entfernt werden sollen, werden aufgrund der Wahl der Volumenverhältnisse bei der SPRI-Bead-Aufreinigung in einem ersten Schritt an die Beads getrieben, die kleineren Fragmente verbleiben in Lösung. Die hierfür verwendeten Volumenverhältnisse müssen den jeweiligen Ampure-Bead-Titrationen entnommen werden. Beispielsweise würde bei der Charge, deren Titrationsergebnisse in Kapitel 3.1 dargestellt sind, 50 µl SPRI-Beads auf 100 µl DNA verwendet werden, da bei diesem Volumenverhältnis 500 bp und 700 bp-lange Fragmente kaum an

die Beads getrieben werden, 1000 bp lange dagegen fast vollständig. Auch hier muss wieder ein Kompromiss eingegangen werden, in Abhängigkeit von der DNA-Menge, die zur Verfügung steht; bei einem kleineren Verhältnis (z.B. 40 µl Beads auf 100 µl DNA) verbleibt auch ein größerer Teil der langen Nebenprodukte in Lösung und damit in der Library. Erhöht man dagegen das Verhältnis (z.B. 55 µl Beads auf 100 µl DNA), so wird auch ein größerer Teil der 600 bp langen KaretiLK-Library an die Beads getrieben und geht verloren. Der Größenunterschluss bei der KaretiLK-Library war erfolgreich; nach dem Größenunterschluss ist auf dem Konzentrationsprofil nur noch der Peak des spezifischen Produktes zu sehen, der Peak des größeren Nebenproduktes ist nicht mehr sichtbar.

emPCR, Beadrecovery und Sequencing. Da letztendlich nur über einen erfolgreichen 454-Sequenzierlauf eine Aussage getroffen werden kann, ob mit dem neuen veränderte Protokoll erfolgreich 454-Librarys hergestellt worden sind, wurden die so hergestellten Librarys in einem emPCR-Cup angesetzt und die erhaltenen Beads auf eine PTP-Region geladen und sequenziert. Der Run produzierte auf dieser Region über 410000 Reads bei einer durchschnittlichen Länge von 380 bp. Damit ist er als erfolgreicher Sequenzier-Lauf zu bewerten, da die Readzahl oberhalb der von Roche vorgegebenen Untergrenzen pro Region liegt (375000 Reads bei Amplicon-Sequenzierung) und auch die durchschnittliche Leselänge relativ gut ist (für Shotgun-Librarys werden Leseweiten für einen guten Run zwischen 350 und 400 bp angegeben, wobei letztere eher selten erreicht werden mit der Titanium-Chemie). Die Ergebnisse der Kontroll-DNA-Reads sind ebenfalls hervorragend, weshalb man davon ausgehen kann, dass es zu keinen Überstrahlungen oder anderen negativen Effekten kam, was sicherlich auch an der etwas geringeren Anzahl verwendeter Beads liegt –das Beadrecovery hatte nicht mehr als die geladenen 1,7 Millionen Beads hervorgegeben. Das ist auch der Grund für die Anzahl an detektierten Raw-Wells, die mit rund 984000 Wells zwar gut, aber sicher noch ausbaufähig nach oben ist und damit auch die erhaltene Readmenge vergrößert werden könnte.

Die guten Run-Ergebnisse zeigen, dass die Verbesserungen des Protokolls funktionieren und sequenzierbare Librarys hervorgebracht haben. Die Inaktivierung von SfiI mittels Proteinase K, deren Inaktivierung und die Ligation der selbsthergestellten 454-Adapter an die geschnittene cDNA ohne Entfernen kleinerer Restriktionsfragmente war erfolgreich. Die durchschnittliche Leselänge von 380 bp und die rechtslastige Größenverteilung der gact-Library-Reads zeigen, dass Größenselektion der Librarys, die ausschließlich während des neuen Library-Protokolls stattgefunden hat, ebenfalls erfolgreich war. Die Verbesserungen des Protokolls führen also zu gut sequenzierbaren Librarys, die aus wesentlich geringeren Eingangsmengen an DNA erhalten werden können.

4.4 Ausblick

Die im Rahmen dieser Dissertation entwickelte Methode zur gerichteten Sequenzierung im Next-Generation Sequencing von cDNA-Banken stellt insbesondere eine Alternative zur bestehenden Amplicon-Sequenzierung auf dem 454 dar.

Mit der Einführung des Titanium plus-Systems werden größere Leseweiten als die bisherigen ermöglicht. Die hier entwickelte Methode müsste an das plus-System angepasst werden. Dabei sind mehrere Faktoren zu beachten. Zum einen müssen bei größeren Insert-Längen die Aufreinigungsschritte entsprechend angepasst werden, um die geforderten Produktgrößen zu selektionieren. Zum anderen findet beim Titanium-plus-System ein verändertes Signal-Processing statt, das an die höhere Flow-Cycle-Zahl angepasst wurde (mündliche Kommunikation mit dem Roche-Support). Vielleicht hat das Konsequenzen für die Qualität von Runs, die mit Librarys der hier entwickelten Methode gestartet werden. Es könnte möglicherweise zu Problemen und Qualitätsverlusten kommen, die von abweichenden Signalintensitäten der Librarys im Vergleich zu denen der Standard Shotgun Librarys verursacht werden. Dem könnte durch veränderte emPCR-Bedingungen entgegengewirkt werden, z.B. durch eine Veränderung der eingesetzten Primer-Mengen, wie es derzeit teilweise bei manchem Amplicon-Protokollen von Roche empfohlen wird; dies alles müsste ausgetestet und adaptiert werden.

Die Vorteile, die diese Methodik gegenüber dem Standard-Protokoll offeriert, könnten auch noch weiter vertieft werden. Ein zentraler Punkt ist die Entkopplung der Template-Generation von der 454-Library-Herstellung. Die damit gewonnene Freiheit bei der Wahl der Sequenzierrichtung wurde in dieser Dissertation auf die Forward-Strang-Bestimmung beschränkt; es könnten jedoch noch entsprechende Versuche für die Revers-Sequenzierung von Librarys durchgeführt werden. Obwohl in dieser Arbeit schon eine Vielzahl von MIDs erfolgreich verwendet wurden, müssten weitere ausgetestet werden. Der durch die Entkopplung ermöglichte höhere Grad der Kombinatorik müsste dementsprechend in 454-Runs getestet werden, um die Belastbarkeit des 454-Systems bei hohem Multiplexgrad zu erproben. Ebenfalls noch nicht getestet ist die Verwendung anderer, kürzerer MID-Sequenzen innerhalb der selbsthergestellten Adapter, um Platz beim Sequenzieren zu sparen und so mehr Basen des Inserts bestimmen zu können.

Durch die Entkopplung konnten Inkompatibilitäten zwischen Target- und 454-Sequenzen beseitigt werden. Da SfiI aufgrund seiner Erkennungssequenz sehr flexibel eingesetzt werden kann, könnten – als Alternative zu den bestehenden – andere SfiI-Schnittstellen zusätzlich erprobt werden. Dadurch könnte die Variabilität bei der Primerwahl weiter erhöht werden, falls es doch einmal zu Problemen in einer PCR bei der Einführung der SfiI-Sites aufgrund von Sequenzähnlichkeiten zur Zielregion kommen sollte. Auch die Verwendung anderer Restriktionsenzyme, die ebenfalls Sticky-Ends produzieren, aber sich hitzeinaktivieren lassen, stellt eine mögliche noch auszutestende Alternative dar, um den Proteinase K-Schritt oder die MinElute-Aufreinigung zu entfernen und so Zeit einzusparen.

Eine weitere Verbesserung für die Effizienz der Libraryherstellung wäre die Erprobung von Fastlink-Ligasen im Protokoll. Diese stellen eine Alternative zu der durchgeführten Ligation dar, die 16 Stunden dauert –sie könnte auf wenige Minuten verkürzt werden. Es müsste ausgetestet und sichergestellt werden, dass die Ligation genauso gut funktioniert wie die bisherige und Blunt-End- oder Einzelstrang-Ligationen zugunsten einer Sticky-End-Ligation deutlich diskriminiert werden.

Um die allgemeine Anwendung der neuen Methodik zur Library-Erstellung für die Hochdurchsatz-Sequenzierung zu testen, wäre es interessant zu überprüfen, ob sich die entwickelte Methode auf andere Sequenziersysteme adaptieren ließe, beispielsweise auf das Solexa- oder SOLID-System, da auch hier die Insert-DNA von Adaptern flankiert wird. Die neue Methodik sollte auch dafür geeignet sein, da durch die Entkopplung prinzipiell jede Sequenz an das Template ligiert werden kann, egal ob es sich dabei um 454, Illumina oder SOLID-Sequenzen handelt.

5. Zusammenfassung – Summary

Mit dem 454 Pyrosequencing als ein Vertreter der Next Generation Sequencing-Plattformen ist es möglich, von einer Probe bis zu einer Millionen Reads innerhalb eines Sequenzierlaufs zu erhalten. Um DNA mittels 454-Technologie sequenzieren zu können, muss sie vorher in eine entsprechende Library umgewandelt werden. Dabei gibt es prinzipiell zwei verschiedene Herangehensweisen. Zum einen werden die für das Sequenzieren benötigten Adaptersequenzen bei der Erstellung einer Shotgun-Library über Random-Ligation an die DNA eingeführt, was zu ungerichteter Sequenzierung der gesamten DNA führt. Zum anderen werden die Adaptersequenzen bei der Erstellung einer Amplicon Library mittels PCR heranamplifiziert, was eine gerichtete Sequenzierung Target-spezifischer Bereiche ermöglicht. Zur Bestimmung ganzer Genome oder Transkriptome wird die Shotgun-Methodik angewendet. Für die selektive Bestimmung spezifischer Bereiche bleibt nur die Möglichkeit einer Amplicon Library-Erstellung. Die dabei ermöglichte gerichtete Sequenzierung ist für verschiedenste Studien unverzichtbar, beispielsweise bei Diversitätsbestimmungen, Mutationsanalysen, Populationsstudien, Analyse von Genexpressionen oder bei problematischen Sequenzbereichen (z.B. homopolymeren Bereichen). Die Herstellung einer 454-Amplicon Library nach dem von Roche vorgegebenen Protokoll hat einige Mängel: (i) die Adaptersequenzen werden als bis zu 40 Basenpaar lange Overhangs an den 5'-Enden der Primer eingeführt; (ii) die verwendeten Primer können Längen von insgesamt über 60 Basenpaare haben, was zu Schwierigkeiten in der PCR führen kann; (iii) die Adaptersequenzen selber können Probleme verursachen, wenn sie z.B. Ähnlichkeiten zu den Target-spezifischen Sequenzen aufweisen.

In dieser Arbeit wurden neue Protokolle für das 454-Sequencing entwickelt, die auch dann erfolgreich anwendbar sind, wenn Library-Erstellung und Sequenzierung über die bestehenden 454-Protokolle scheitern. Die in dieser Arbeit entwickelte Methodik beruht auf gerichteter Ligation von 454-Adaptersequenzen über die Verwendung von unterschiedlichen SfiI-Restriktions-Schnittstellen. Diese können schon während der cDNA-Synthese in das Template eingeführt, oder unabhängig davon mittels PCR als kurze Overhangs heranamplifiziert werden. Des Weiteren wurde die Template- und 454-Library-Produktion entkoppelt. Im Gegensatz zu den bestehenden Protokollen ist man bei der 454-Library-Herstellung weitestgehend unabhängig von Template-Sequenzen und verwendeten Primern. Die Wahl der MIDs für die Gestaltung von Multiplex-Runs ist flexibler und wird erleichtert, sie kann zeitnah durchgeführt werden; Engpässe werden so verringert. Die Sequenzierrichtung kann frei gewählt werden, womit eine optimale Anpassung an biologische Fragestellungen und effektive Ausnutzung bei der PTP-Beladung ermöglicht wird. Das Protokoll ist leicht adaptierbar und funktioniert prinzipiell mit jeglichem Template. Aufreinigungs und Größenselektionsschritte bei der Libraryproduktion garantieren eine hohe Qualität. Durch Optimierungen konnten die benötigten Template-Mengen gemessen an den Standard-Protokollen auf weniger als ein Zehntel reduziert werden, was das Sequencing von gering konzentrierten Proben erst ermöglicht -rund 10 ng Template-DNA sind ausreichend für die Library-Präparation.

Es wurden mit der im Rahmen dieser Arbeit entwickelten alternativen Methode erfolgreiche 454-Sequenzier-Läufe von cDNAs und einer Illumina-Library durchgeführt, außerdem wurde ein Deepsequencing von IgH-cDNAs ermöglicht.

Summary

With 454-pyrosequencing as one of the major Next-Generation-Sequencing-platforms it is possible to obtain up to one million of reads from one sample during a single sequencing-run. To be able to sequence a DNA-sample, it has to be converted into an appropriate library. There are two principal ways for this. On the one hand, adaptor-sequences necessary for 454-sequencing are added to the template DNA during a shotgun-library-generation by random ligation, resulting in undirected sequencing of the whole DNA. On the other hand, the adapter-sequences are added by PCR during an amplicon library-preparation resulting in directed sequencing of target-specific areas. For the determination of whole genomes or transcriptomes, the shotgun approach is appropriate. If target-specific sequencing is required, preparation of an amplicon library is the method of choice. Directed sequencing enabled by amplicon library-preparation is essential for various studies focusing on, e.g. diversity-analysis of combinatorial libraries, mutation detection, population-studies, analysis of gene-expression or sequencing of problematic targets. The amplicon library-generation following Roche's protocol has some flaws: (i) the adapter-sequences are added as up to 40 base-pair long overhangs at the 5'-ends of primers used during PCR; (ii) the primers have lengths of over 60 base-pairs overall, which can cause trouble during PCR and (iii) the adaptor-sequences themselves may be problematic due to possible sequence-similarities with the target specific sequences.

The outcome of this work are novel protocols for 454-sequencing, which can be successfully applied to templates where sequencing utilizing existing 454-protocols failed. The method developed is based on directed ligation of 454 adapter-sequences by using different SfiI-restriction-sites of the template-DNA. These can be introduced into the template during cDNA-synthesis, or independently as short overhangs by PCR. Additionally, template- and 454-library-production-steps are uncoupled. In contrast to existing protocols, 454-library-production becomes independent of template-sequences and – primers. For the design of multiplex-runs, the assignment of MID's becomes more flexible, allowing contemporary decisions, hence, avoiding bottlenecks. Orientation of sequencing direction can be chosen freely, offering optimal adaptation to any biological question and an effective exploitation of PTP-space. The protocol is easily adaptable and works with any template in principle. Purification- and size-selection-steps are performed during library-preparation to guarantee high quality of the library-DNA. Through optimization of the developed protocol, the amount of template needed for library preparation could be reduced to less than a tenth compared to standard protocols, making sequencing of low-concentrated samples possible – an amount of 10 ng template-DNA is sufficient for library-preparation.

In the context of this work successful 454-sequencing-runs of cDNA-libraries and an Illumina-library were performed, and also deepsequencing of IgH-cDNAs was made possible.

6. Literatur

- [1] A. E. Mirsky, „The discovery of DNA,“ *Sci Am.*, Bd. 218, Nr. 6, pp. 78-88, 1968.
- [2] O. T. Avery, C. M. MacLeod und M. McCarty, „Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a deoxyribonucleic acid fraction isolated from *Pneumococcus*,“ *J. Exp. Med.*, pp. 137-158, 1944.
- [3] R. S. Tipson, „Phoebus Aaron Theodor Levene, 1869-1940,“ *Adv Carbohydr Chem.*, Bd. 12, pp. 1-12, 1957.
- [4] A. R. Todd, „The nucleotides: some recent chemical research and its biological implications,“ *Harvey Lect.*, Bd. 47, pp. 1-20, 1951-1952.
- [5] E. Chargaff, „Some recent studies on the composition and structure of nucleic acids,“ *J Cell Physiol Suppl*, Bd. 38, pp. 41-59, 1951.
- [6] J. D. Watson und F. H. Crick, „Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid,“ *Nature.*, Bd. 171, Nr. 4356, pp. 737, 738, 1953.
- [7] R. E. Franklin und R. G. Gosling, „Molecular configuration in sodium thymonucleate,“ *Nature.*, Bd. 171, Nr. 4356, pp. 740, 741, 1953.
- [8] M. H. Wilkins, A. R. Stokes und H. R. Wilson, „Molecular structure of deoxypentose nucleic acids,“ *Nature.*, Bd. 171, Nr. 4356, pp. 738-740, 1953.
- [9] F. H. Crick, „On protein synthesis,“ *Symp Soc Exp Biol.*, Bd. 12, pp. 138-163, 1958.
- [10] M. W. Nirenberg, „The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides,“ *Proc Natl Acad Sci U S A.*, Bd. 47, pp. 1588-1602, 1961.
- [11] A. M. Maxam und W. Gilbert, „A new method for sequencing DNA,“ *Proc Natl Acad Sci U S A.*, Bd. 74, Nr. 2, pp. 560-564, 1977.
- [12] D. Voet und J. G. Voet, *Biochemie*, VCH, 1994, pp. 834-838.
- [13] F. Sanger, S. Nicklen und A. R. Coulson, „DNA sequencing with chain-terminating inhibitors,“ *Proc Natl Acad Sci U S A.*, Bd. 74, Nr. 12, pp. 5463-5467, 1977.
- [14] F. Sanger und A. R. Coulson, „A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase,“ *J Mol Biol.*, Bd. 94, Nr. 3, pp. 441-448, 1975.
- [15] K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn und H. Erlich, „Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction,“ *Cold Spring Harb Symp Quant Biol.*, Bd. 51, pp. 263-273, 1986.
- [16] L. M. Smith, J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. Kent und L. E. Hood, „Fluorescence detection in automated DNA sequence analysis,“ *Nature.*, Bd. 321, Nr. 6071, pp. 674-679, 1986.
- [17] L. M. Smith, S. Fung, M. W. Hunkapiller, T. J. Hunkapiller und L. E. Hood, „The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis,“ *Nucleic Acids Res.*, Bd. 13, Nr. 7, pp. 2399-2412, 1985.

- [18] J. M. Prober, G. L. Trainor, R. J. Dam, F. W. Hobbs, C. W. Robertson, R. J. Zargursky, A. J. Cocuzza, M. A. Jensen und K. Baumeister, „A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides,“ *Science.*, Bd. 238, Nr. 4825, pp. 336-341, 1987.
- [19] H. Swerdlow, S. L. Wu, H. Harke und N. J. Dovichi, „Capillary gel electrophoresis for DNA sequencing. Laser-induced fluorescence detection with the sheath flow cuvette,“ *J Chromatogr.*, Bd. 516, Nr. 1, pp. 61-67, 1990.
- [20] T. Hunkapiller, R. J. Kaiser, B. F. Koop und L. Hood, „Large-scale and automated DNA sequence determination,“ *Science.*, Bd. 254, Nr. 5028, pp. 59-67, 1991.
- [21] S. Anderson, „Shotgun DNA sequencing using cloned DNase I-generated fragments,“ *Nucleic Acids Res.*, Bd. 9, Nr. 13, pp. 3015-3027, 1981.
- [22] H. Rauth, *Verfahrenstechnik und Automatisierung nicht radioaktiver Highthroughput DNA-Sequenzierung*, 1997.
- [23] H. Kuhl, *Ein Verfahren für BAC DNA-Aufreinigung im Hochdurchsatz zur Genomkartierung von *Dicentrarchus labrax**, Technische Universität Berlin, 2008.
- [24] R. D. Fleischmann, M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty und J. M. Merrick, „Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd,“ *Science.*, Bd. 269, Nr. 5223, pp. 496-512, 1995.
- [25] International Human Genome Sequencing Consortium, „Initial sequencing and analysis of the human genome,“ *Nature.*, Bd. 409, Nr. 6822, pp. 860-921, 2001.
- [26] J. C. Venter, „The sequence of the human genome,“ *Science.*, Bd. 291, Nr. 5507, pp. 1304-1351, 2001.
- [27] M. Ronaghi, M. Uhlén und P. Nyérén, „A sequencing method based on real-time pyrophosphate detection,“ *Science.*, Bd. 281, Nr. 5375, p. 363–365, 1998.
- [28] R. J. Melamede. USA Patent 4863849, 1985.
- [29] P. Nyérén und A. Lundin, „Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis,“ *Anal Biochem.*, Bd. 151, Nr. 2, pp. 504-509, 1985.
- [30] P. Nyérén, B. F. Nore und M. Baltscheffsky, „Studies on photosynthetic inorganic pyrophosphate formation in *Rhodospirillum rubrum* chromatophores,“ *Biochim Biophys Acta.*, Bd. 851, pp. 276-282, 1986.
- [31] P. Nyérén, B. F. Nore und M. Baltscheffsky, „Inorganic pyrophosphate synthesis after a short light flash in chromatophores from *Rhodospirillum rubrum*,“ *Photobiochem Photobiophys.*, pp. 189-196, 1986.
- [32] P. Nyérén, „Enzymatic method for continuous monitoring of DNA polymerase activity,“ *Anal Biochem.*, Bd. 167, Nr. 2, pp. 235-238, 1987.
- [33] P. Nyérén, „Apyrase immobilized on paramagnetic beads used to improve detection limits in bioluminometric ATP monitoring,“ *J Biolumin Chemilumin.*, Bd. 9, Nr. 1, pp. 29-34, 1994.
- [34] S. Stahl, T. Hultman, T. Moks und M. Uhlén, „Solid phase DNA sequencing using the biotin-avidin system,“ *Nucleic Acids Res.*, Bd. 16, Nr. 7, pp. 3025-3038, 1988.
- [35] P. Nyérén, B. Pettersson und M. Uhlén, „Solid phase DNA minisequencing by an enzymatic luminometric inorganic

- pyrophosphate detection assay," *Anal Biochem.*, Bd. 208, Nr. 1, pp. 171-175, 1993.
- [36] M. Ronaghi, S. Karamohamed, B. Pettersson, M. Uhlén und P. Nyrén, „Real-time DNA sequencing using detection of pyrophosphate release," *Anal Biochem.*, Bd. 242, Nr. 1, pp. 84-89, 1996.
- [37] S. Karamohamed, J. Nilsson, K. Nourizad, M. Ronaghi, B. Pettersson und P. Nyrén, „Production, purification, and luminometric analysis of recombinant *Saccharomyces cerevisiae* MET3 adenosine triphosphate sulfurylase expressed in *Escherichia coli*," *Protein Expr Purif.*, Bd. 15, Nr. 3, pp. 381-388, 1999.
- [38] S. Karamohamed, T. Nordström und P. Nyrén, „Real-time bioluminometric method for detection of nucleoside diphosphate kinase activity," *Biotechniques.*, Bd. 26, Nr. 4, pp. 728-734, 1999.
- [39] P. Nyrén. USA and worldwide Patent US 6 258 568B1 and WO98/28440, 2001.
- [40] M. Ronaghi, „Improved performance of Pyrosequencing using singlestranded DNA-binding protein," *Anal Biochem.*, Bd. 286, Nr. 2, p. 282-288, 2000.
- [41] T. Nordström, B. Gharizadeh, N. Pourmand, P. Nyrén und M. Ronaghi, „Method enabling fast partial sequencing of cDNA clones," *Anal Biochem.*, Bd. 292, Nr. 2, p. 266-271, 2001.
- [42] T. Nordström, M. Ronaghi und P. Nyrén, „Automation of a novel DNA sequencing method," in *Bioluminescence and Chemiluminescence: Perspective for the 21st Century*, Hoboken, NJ, John Wiley and Sons, 1999, p. 528-531.
- [43] A. Ahmadian, B. Gharizadeh, A. C. Gustafsson, F. Sterky, P. Nyrén, M. Uhlén und J. Lundeberg, „Single nucleotide polymorphism analysis by Pyrosequencing," *Anal Biochem.* 280, pp. 103-110, 2000.
- [44] J. D. Gruber, P. B. Colligan und J. K. Wolford, „Estimation of single nucleotide polymorphism allele frequency in DNA pools by using Pyrosequencing," *Hum Genet.*, Bd. 110, Nr. 5, pp. 395-401, 2002.
- [45] K. Uhlmann, A. Brinckmann, M. R. Toliat, H. Ritter und P. Nürnberg, „Evaluation of a potential epigenetic biomarker by quantitative methyl-single nucleotide polymorphism analysis," *Electrophoresis.*, Bd. 23, Nr. 24, pp. 4072-4079, 2002.
- [46] A. Goriely, G. A. McVean, M. Röjmyr, B. Ingemarsson und O. A. Wilkie, „Evidence for selective advantage of pathogenic FGFR2 mutations in the male germ line," *Science.*, Bd. 301, Nr. 5633, pp. 643-646, 2003.
- [47] B. Gharizadeh, A. Ohlin, P. Mölling, A. Bäckman, B. Amini, P. Olcén und P. Nyrén, „Multiple group-specific sequencing primers for reliable and rapid DNA sequencing," *Mol Cell Probes.*, Bd. 17, Nr. 4, pp. 203-210, 2003.
- [48] B. Gharizadeh, E. Norberg, J. Löffler, S. Jalal, J. Tollemar, H. Einsele, L. Klingspor und P. Nyrén, „Identification of medically important fungi by Pyrosequencing technology," *Mycoses.*, Bd. 47, Nr. 1-2, pp. 29-33, 2004.
- [49] B. Gharizadeh, M. Kalantari, C. A. Garcia, B. Johansson und P. Nyrén, „Typing of human papillomavirus by pyrosequencing," *Lab Invest.*, Bd. 81, Nr. 5, pp. 673-679, 2001.
- [50] H. Andréasson, A. Asp, A. Alderborn, U. Gyllensten und M. Allen, „Mitochondrial sequence analysis for forensic identification using pyrosequencing technology," *Biotechniques.*, Bd. 32, Nr. 1, pp. 130-133, 2002.
- [51] H. Fakhrai-Rad, N. Pourmand und M. Ronaghi, „Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms," *Hum Mutat.*, Bd. 19, Nr. 5, pp. 479-485, 2002.
- [52] B. Gharizadeh, T. Nordström, A. Ahmadian, M. Ronaghi und P. Nyrén, „Long-read pyrosequencing using pure 2'-

- deoxyadenosine-5'-O'-(1-thiotriphosphate) Sp-isomer," *Anal Biochem.*, Bd. 301, Nr. 1, pp. 82-90, 2002.
- [53] J. Eriksson, B. Gharizadeh, N. Nourizad und P. Nyrén, „7-Deaza-2'-deoxyadenosine-5'-triphosphate as an alternative nucleotide for the pyrosequencing technology," *Nucleosides Nucleotides Nucleic Acids.*, Bd. 23, Nr. 10, pp. 1583-1594, 2004.
- [54] S. Tabor und C. C. Richardso, „DNA sequence analysis with a modified bacteriophage T7 DNA polymerase," *Proc Natl Acad Sci U S A.*, Bd. 84, Nr. 14, pp. 4767-4771, 1987.
- [55] B. Gharizadeh, J. Eriksson, N. Nourizad, T. Nordström und P. Nyrén, „Improvements in Pyrosequencing technology by employing Sequenase polymerase," *Anal Biochem.*, Bd. 330, Nr. 2, pp. 272-280, 2004.
- [56] J. Eriksson, T. Nordström und P. Nyrén, „Method enabling firefly luciferase based bioluminometric assays at elevated temperature," *Anal Biochem.*, Bd. 314, Nr. 1, pp. 158-161, 2003.
- [57] J. Eriksson, B. Gharizadeh, T. Nordström und P. Nyrén, „Pyrosequencing trade mark technology at elevated temperature," *Electrophoresis.*, Bd. 25, Nr. 1, pp. 20-27, 2004.
- [58] T. Nordström, K. Nourizad, M. Ronaghi und P. Nyrén, „Method enabling pyrosequencing on double-stranded DNA," *Anal Biochem.*, Bd. 282, Nr. 2, pp. 186-193, 2000.
- [59] T. Nordström, A. Alderborn und P. Nyrén, „Method for one-step preparation of double-stranded DNA template applicable for use with Pyrosequencing technology," *J Biochem Biophys Methods.*, Bd. 52, Nr. 2, pp. 71-82, 2002.
- [60] B. Gharizadeh, M. Ghaderi, D. Donnelly, B. Amini, K. L. Wallin und P. Nyrén, „Multiple-primer DNA sequencing method," *Electrophoresis.*, Bd. 24, Nr. 7-8, pp. 1145-1151, 2003.
- [61] B. Gharizadeh, M. Oggionni, B. Zheng, E. Akom, N. Pourmand, A. Ahmadian, K. L. Wallin und P. Nyrén, „Type-specific multiple sequencing primers: a novel strategy for reliable and rapid genotyping of human papillomaviruses by pyrosequencing technology," *J Mol Diagn.*, Bd. 7, Nr. 2, pp. 198-205, 2005.
- [62] M. Margulies, M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley und J. M. Rothberg, „Genome sequencing in microfabricated high-density picolitre reactors," *Nature.*, Bd. 437, Nr. 7057, pp. 376-380, 2005.
- [63] F. J. Steemers und D. R. Walt, „Multi-analyte sensing: from site-selective deposition to randomly-ordered addressable optical fiber sensors," *Mikrochim Acta.*, Bd. 131, p. 99-105, 1999.
- [64] P. Pantano und D. R. Walt, „Ordered nanowell arrays," *Chem Mater.*, Bd. 8, p. 2832-2835, 1996.
- [65] J. A. Ferguson, F. J. Steemers und D. R. Walt, „High-density fiber-optic DNA random microsphere array," *Anal Chem.*, Bd. 72, Nr. 22, pp. 5618-5624, 2000.
- [66] J. H. Leamon, W. L. Lee, K. R. Tartaro, J. R. Lanza, G. J. Sarkis, A. D. deWinter, J. Berka, M. Weiner, J. M. Rothberg und K. L. Lohman, „A massively parallel PicoTiterPlate based platform for discrete picoliter-scale polymerase chain reactions," *Electrophoresis.*, Bd. 24, Nr. 21, pp. 3769-3777, 2003.

- [67] D. S. Tawfik und A. D. Griffiths, „Man-made cell-like compartments for molecular evolution,“ *Nat Biotechnol.*, Bd. 16, Nr. 7, pp. 652-656, 1998.
- [68] F. J. Ghadessy, J. L. Ong und P. Holliger, „Directed evolution of polymerase function by compartmentalized self-replication,“ *Proc Natl Acad Sci U S A.*, Bd. 98, Nr. 8, pp. 4552-4557, 2001.
- [69] J. M. Rothberg und J. H. Leamon, „The development and impact of 454 sequencing,“ *Nat Biotechnol.*, Bd. 26, Nr. 10, pp. 1117-1124, 2008.
- [70] R. Williams, S. G. Peisajovich, O. J. Mille, S. Magdassi, D. S. Tawfik und A. D. Griffiths, „Amplification of complex gene libraries by emulsion PCR,“ *Nat Methods.*, Bd. 3, Nr. 7, pp. 545-550, 2006.
- [71] S. Kwok, D. E. Kellogg, N. McKinney, D. Spasic, L. Goda, C. Levenson und J. J. Sninsky, „Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies,“ *Nucleic Acids Res.*, Bd. 18, Nr. 4, pp. 999-1005, 1990.
- [72] Y. Benita, R. S. Oosting, M. C. Lok, M. J. Wise und I. Humphery-Smith, „Regionalized GC content of template DNA as a predictor of PCR success,“ *Nucleic Acids Res.*, Bd. 31, Nr. 19, p. e99, 2003.
- [73] roche454, „GS FLX Titanium emPCR Method Manual,“ Oktober 2008.
- [74] roche454, „emPCR Method Manual - Lib-L LV,“ Oktober 2009.
- [75] roche454, „Sequencing Method Manual,“ Oktober 2009.
- [76] J. M. Aliotta, J. J. Pelletier, J. L. Ware, L. S. Moran, J. S. Benner und H. Kong, „Thermostable Bst DNA polymerase I lacks a 3'→5' proofreading exonuclease activity,“ *Genet Anal.*, Bd. 12, Nr. 5-6, pp. 185-195, 1996.
- [77] D. C. Thomas, G. A. Nardone und S. K. Randall, „Amplification of padlock probes for DNA diagnostics by cascade rolling circle amplification or the polymerase chain reaction,“ *Arch Pathol Lab Med.*, Bd. 123, Nr. 12, pp. 1170-1176, 1999.
- [78] J. McClary, S. Y. Ye, G. F. Hong und F. Witney, „Sequencing with the large fragment of DNA polymerase I from *Bacillus stearothermophilus*,“ *DNA Seq.*, Bd. 1, Nr. 3, pp. 173-180, 1991.
- [79] D. A. Mead, J. A. McClary, J. A. Luckey, A. J. Kostichka, F. R. Witney und L. M. Smith, „Bst DNA polymerase permits rapid sequence analysis from nanogram amounts of template,“ *Biotechniques.*, Bd. 11, Nr. 1, pp. 76-78, 80, 82-87, 1991.
- [80] P. M. Lizardi, X. Huang, Z. Zhu, P. Bray-Ward, D. C. Thomas und D. C. Ward, „Mutation detection and single-molecule counting using isothermal rolling-circle amplification,“ *Nat Genet.*, Bd. 19, Nr. 3, pp. 225-232, 1998.
- [81] M. Ronaghi, „Pyrosequencing sheds light on DNA sequencing,“ *Genome Res.*, Bd. 11, Nr. 1, pp. 3-11, 2001.
- [82] roche454, „Genome Sequencer FLX System Software Manual, version 2.3 Part B: GS Run Processor – GS Reporter – GS Run Browser – GS Support Tool,“ Oktober 2009.
- [83] roche454, „TCB No.004-2009,“ April 2009.
- [84] roche454, „TCB No.005-2009,“ April 2009.
- [85] roche454, „GS FLX Titanium General Library Preparation Method Manual,“ Oktober 2008.

- [86] J. O. Korbel, A. E. Urban, J. P. Affourtit, B. Godwin, F. Grubert, J. F. Simons, P. M. Kim, D. Palejev, N. J. Carriero, L. Du, B. E. Taillon, Z. Chen, A. Tanzer, A. C. Saunders, J. Chi, F. Yang, N. P. Carter, M. E. Hurles, S. M. Weissman, T. T. Harkins, M. B. Gerstein, M. Egholm und M. Snyder, „Paired-end mapping reveals extensive structural variation in the human genome,“ *Science*, Bd. 318, Nr. 5849, pp. 420-426, 2007.
- [87] roche454, „GS FLX Titanium 3 kb Span Paired End Library Preparation Method Manual,“ April 2009.
- [88] roche454, „Paired End Library Preparation Method Manual - 3 kb Span,“ Oktober 2009.
- [89] roche454, „Paired End Library Preparation Method Manual - 20 kb and 8 kb Span,“ Oktober 2009.
- [90] roche454, „Rapid Library Preparation Method Manual,“ Oktober 2009.
- [91] roche454, cDNA Rapid Library Preparation Method Manual, Oktober 2009.
- [92] roche454, „Paired End Rapid Library Preparation Method Manual, 20 kb and 8 kb Span,“ April 2012.
- [93] roche454, „Paired End Rapid Library Preparation Method Manual, 3 kb Span,“ April 2012.
- [94] roche454, „454 Sequencing System Guidelines for Amplicon Experimental Design,“ Juli 2011.
- [95] I. Braslavsky, B. Hebert, E. Kartalov und S. R. Quake, „Sequence information can be obtained from single DNA molecules,“ *Proc Natl Acad Sci U S A*, Bd. 100, Nr. 7, pp. 3960-3964, 2003.
- [96] T. D. Harris, P. R. Buzby, H. Babcock, E. Beer, J. Bowers, I. Braslavsky, M. Causey, J. Colonell, J. Dimeo, J. W. Efcavitch, E. Giladi, J. Gill, J. Healy, M. Jarosz, D. Lapen, K. Moulton, S. R. Quake, K. Steinmann, E. Thayer, A. Tyurina, R. Ward, H. Weiss und Z. Xie, „Single-molecule DNA sequencing of a viral genome,“ *Science*, Bd. 320, Nr. 5872, pp. 106-109, 2008.
- [97] „tSMS™ Performance,“ [Online]. Available: <http://www.helicosbio.com/Technology/TrueSingleMoleculeSequencing/tSMStradePerformance/tabid/151/Default.aspx>. [Zugriff am 23 07 2012].
- [98] M. Fedurco, A. Romieu, S. Williams, I. Lawrence und G. Turcatti, „BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies,“ *Nucleic Acids Res.*, Bd. 34, Nr. 3, p. e22, 2006.
- [99] G. Turcatti, A. Romieu, M. Fedurco und A. P. Tairi, „A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis,“ *Nucleic Acids Res.*, Bd. 36, Nr. 4, p. e25, 2008.
- [100] C. Adessi, G. Matton, G. Ayala, G. Turcatti, J. J. Mermod, P. Mayer und E. Kawashima, „Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms,“ *Nucleic Acids Res.* 28, p. E87, 2000.
- [101] „Illumina - Sequencing Portfolio,“ [Online]. Available: <http://www.illumina.com/systems/sequencing.ilmn>. [Zugriff am 23 07 2012].
- [102] J. C. Dohm, C. Lottaz, T. Borodina und H. Himmelbauer, „Substantial biases in ultra-short read data sets from high-throughput DNA sequencing,“ *Nucleic Acids Res.*, Bd. 36, Nr. 16, p. e105, 2008.
- [103] L. W. Hillier, G. T. Marth, A. R. Quinlan, D. Dooling, G. Fewell, D. Barnett, P. Fox, J. I. Glasscock, M. Hickenbotham, W. Huang, V. J. Magrini, R. J. Richt, S. N. Sander, D. A. Stewart, M. Stromberg, E. F. Tsung, T. Wylie, T. Schedl, R. K. Wilson und E. R. Mardis, „Whole-genome sequencing and variant discovery in *C. elegans*,“ *Nat Methods*, Bd. 5, Nr. 2, pp.

- 183-188, 2008.
- [104] O. Harismendy, P. C. Ng, R. L. Strausberg, X. Wang, T. B. Stockwell, K. Y. Beeson, N. J. Schork, S. S. Murray, E. J. Topol, S. Levy und K. A. Frazer, „Evaluation of next generation sequencing platforms for population targeted sequencing studies,“ *Genome Biol.*, Bd. 10, Nr. 3, p. R32, 2009.
- [105] J. Shendure, G. J. Porreca, N. B. Reppas, X. Lin, J. P. McCutcheon, A. M. Rosenbaum, M. D. Wang, K. Zhang, R. D. Mitra und G. M. Church, „Accurate multiplex polony sequencing of an evolved bacterial genome,“ *Science.*, Bd. 309, Nr. 5741, pp. 1728-1732, 2005.
- [106] K. J. McKernan, H. E. Peckham, G. L. Costa, S. F. McLaughlin, Y. Fu, E. F. Tsung, C. R. Clouser, C. Duncan, J. K. Ichikawa, C. C. Lee, Z. Zhang, S. S. Ranade, E. T. Dimalanta, F. C. Hyland, T. D. Sokolsky, L. Zhang, A. Sheridan, H. Fu, C. L. Hendrickson, B. Li, L. Kotler, J. R. Stuart, J. A. Malek, J. M. Manning, A. A. Antipova, D. S. Perez, M. P. Moore, K. C. Hayashibara, M. R. Lyons, R. E. Beaudoin, B. E. Coleman, M. W. Laptewicz, A. E. Sannicandro, M. D. Rhodes, R. K. Gottimukkala, S. Yang, V. Bafna, A. Bashir, A. MacBride, C. Alkan, J. M. Kidd, E. E. Eichler, M. G. Reese, F. M. De La Vega und A. P. Blanchard, „Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding,“ *Genome Res.*, Bd. 19, Nr. 9, pp. 1527-1541, 2009.
- [107] „life technologies - 5500 Series Genetic Analysis Systems,“ [Online]. Available: <http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing/next-generation-systems.html>. [Zugriff am 23 07 2012].
- [108] M. J. Levene, J. Korlach, S. W. Turner, M. Foquet, H. G. Craighead und W. W. Webb, „Zero-mode waveguides for single-molecule analysis at high concentrations,“ *Science.*, Bd. 299, Nr. 5607, pp. 682-686, 2003.
- [109] J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. Dewinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach und S. Turner, „Real-time DNA sequencing from single polymerase molecules,“ *Science.*, Bd. 323, Nr. 5910, pp. 133-138, 2009.
- [110] J. M. Rothberg, W. Hinz, T. M. Rearick, J. Schultz, W. Mileski, M. Davey, J. H. Leamon, K. Johnson, M. J. Milgrew, M. Edwards, J. Hoon, J. F. Simons, D. Marran, J. W. Myers, J. F. Davidson, A. Branting, J. R. Nobile, B. P. Puc, D. Light, T. A. Clark, M. Huber, J. T. Branciforte, I. B. Stoner, S. E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R. Kasinskas, T. Sokolsky, J. A. Fidanza, E. Namsaraev, K. J. McKernan, A. Williams, G. T. Roth und J. Bustillo, „An integrated semiconductor device enabling non-optical genome sequencing,“ *Nature.*, Bd. 475, Nr. 7356, pp. 348-352, 2011.
- [111] M. L. Metzker, „Sequencing technologies - the next generation,“ *Nat Rev Genet.*, Bd. 11, Nr. 1, pp. 31-46, 2010.
- [112] T. L. Hawkins, T. O'Connor-Morin, A. Roy und C. Santillan, „DNA purification and isolation using a solid-phase,“ *Nucleic Acids Res.*, Bd. 22, Nr. 21, pp. 4543-4544, 1994.
- [113] M. M. DeAngelis, D. Wang und T. L. Hawkins, „Solid-phase reversible immobilization for the isolation of PCR products,“ *Nucleic Acids Res.*, Bd. 23, Nr. 22, pp. 4742-4743, 1995.
- [114] McKernan und Gustafson, „Chapter 9. Solid Phase Nucleic Acid Purification Using Magnetic Beads,“ in *DNA Sequencing II: Optimizing the Isolation, Preparation and Clean-up*, Jones & Bartlett Pub; 1 edition, 2006, pp. 132-137.
- [115] Z. Wang, M. Gerstein und M. Snyder, „RNA-Seq: a revolutionary tool for transcriptomics,“ *Nat Rev Genet.*, Bd. 10, Nr. 1,

- pp. 57-63, 2009.
- [116] D. G. Schatz und Y. Ji, „Recombination centres and the orchestration of V(D)J recombination,“ *Nat Rev Immunol.*, Bd. 11, Nr. 4, pp. 251-263, 2011.
- [117] Z. Zhu und D. S. Dimitrov, „Construction of a large naïve human phage-displayed Fab library through one-step cloning,“ *Methods Mol Biol.*, Bd. 525, pp. 129-142, 2009.
- [118] P. Prabakaran, W. Chen, M. G. Singarayan, C. C. Stewart, E. Streaker, Y. Feng und D. S. Dimitrov, „Expressed antibody repertoires in human cord blood cells: 454 sequencing and IMGT/HighV-QUEST analysis of germline gene usage, junctional diversity, and somatic mutations,“ *Immunogenetics.*, Bd. 64, Nr. 4, pp. 337-350, 2012.
- [119] T. S. Lim, S. Mollova, F. Rubelt, V. Sievert, S. Dübel, H. Lehrach und Z. Konthur, „V-gene amplification revisited - An optimised procedure for amplification of rearranged human antibody genes of different isotypes,“ *N Biotechnol.*, Bd. 27, Nr. 2, pp. 108-117, 2010.
- [120] D. Sblattero und A. Bradbury, „A definitive set of oligonucleotide primers for amplifying human V regions,“ *Immunotechnology.*, Bd. 3, Nr. 4, pp. 271-278, 1998.
- [121] T. Schütze, F. Rubelt, J. Repkow, N. Greiner, V. A. Erdmann, H. Lehrach, Z. Konthur und J. Glökler, „A streamlined protocol for emulsion polymerase chain reaction and subsequent purification,“ *Anal Biochem.*, Bd. 410, Nr. 1, pp. 155-157, 2011.
- [122] roche454, „emPCR Method Manual - Lib-L SV,“ 2009.
- [123] R. Weßling, S. M. Schmidt, C. O. Micali, F. Knaust, R. Reinhardt, U. Neuman, E. Ver Loren van Themaat und R. Panstruga, „Transcriptome analysis of enriched *Golovinomyces orontii* haustoria by deep 454 pyrosequencing,“ *Fungal Genet Biol.*, Bd. 49, Nr. 6, pp. 470-482, 2012.
- [124] H. Kuhl, M. Tine, J. Hecht, F. Knaust und R. Reinhardt, „Analysis of single nucleotide polymorphisms in three chromosomes of European sea bass *Dicentrarchus labrax*,“ *Comp Biochem Physiol Part D Genomics Proteomics.*, Bd. 6, Nr. 1, pp. 70-75, 2011.
- [125] M. Kube, A. M. Migdoll, I. Gehring, K. Heitmann, Y. Mayer, H. Kuhl, F. Knaust, K. Geider und R. Reinhardt, „Genome comparison of the epiphytic bacteria *Erwinia billingiae* and *E. tasmaniensis* with the pear pathogen *E. pyrifoliae*,“ *BMC Genomics.*, Bd. 11, p. 393, 2010.
- [126] N. E. Murray, S. A. Bruce und K. Murray, „Molecular cloning of the DNA ligase gene from bacteriophage T4. II. Amplification and preparation of the gene product,“ *J Mol Biol.*, Bd. 132, Nr. 3, pp. 493-505, 1979.
- [127] D. R. Horspool, R. J. Coope und R. A. Holt, „Efficient assembly of very short oligonucleotides using T4 DNA Ligase,“ *BMC Res Notes.*, Bd. 3, p. 291, 2010.
- [128] B. Q. Qiang und I. Schildkraut, „A type II restriction endonuclease with an eight nucleotide specificity from *Streptomyces fimbriatus*,“ *Nucleic Acids Res.*, Bd. 12, Nr. 11, pp. 4507-4516, 1984.
- [129] L. M. Wentzell, T. J. Nobbs und S. E. Halford, „The *SfiI* restriction endonuclease makes a four-strand DNA break at two copies of its recognition sequence,“ *J Mol Biol.*, Bd. 248, Nr. 3, pp. 581-595, 1995.
- [130] T. J. Nobbs und S. E. Halford, „DNA cleavage at two recognition sites by the *SfiI* restriction endonuclease: salt dependence of cis and trans interactions between distant DNA sites,“ *J Mol Biol.*, Bd. 252, Nr. 4, pp. 399-411, 1995.

- [131] M. D. Szczelkun und S. E. Halford, „Recombination by resolvase to analyse DNA communications by the SfiI restriction endonuclease,“ *EMBO J.*, Bd. 15, Nr. 6, pp. 1460-1469, 1996.
- [132] L. M. Wentzell und S. E. Halford, „DNA looping by the Sfi I restriction endonuclease,“ *J Mol Biol.*, Bd. 281, Nr. 3, pp. 433-444, 1998.
- [133] M. A. Watson, D. M. Gowers und S. E. Halford, „Alternative geometries of DNA looping: an analysis using the SfiI endonuclease,“ *J Mol Biol.*, Bd. 298, Nr. 3, pp. 461-475, 2000.
- [134] Qiagen, „MinElute Handbook,“ 2008.
- [135] W. Ebeling, N. Hennrich, M. Klockow, H. Metz, H. D. Orth und H. Lang, „Proteinase K from *Tritirachium album* Limber,“ *Eur J Biochem.*, Bd. 47, Nr. 1, pp. 91-97, 1974.
- [136] E. Kraus und U. Femfert, „Proteinase K from the mold *Tritirachium album* Limber. Specificity and mode of action,“ *Hoppe Seylers Z Physiol Chem.*, Bd. 357, Nr. 7, pp. 937-947, 1976.
- [137] "AquaFirst," [Online]. Available: <http://www.sigenae.org/aquafirst/index.php?id=17>. [Zugriff am 23.11.2012].

